

## **Supporting Information**

### **Total Biosynthesis for Milligram-Scale Production of Etoposide Intermediates in a Plant Chassis**

Bailey J. Schultz,<sup>†‡</sup> Seung Yeon Kim,<sup>†</sup> Warren Lau,<sup>†</sup> and Elizabeth S. Sattely<sup>\*,†,⊥</sup>

<sup>†</sup>Department of Chemical Engineering and <sup>⊥</sup>Howard Hughes Medical Institute, Stanford University, Stanford, California 94305, United States

\*Correspondence to: sattely@stanford.edu

### **Table of Contents**

<b>1. Supplementary discussion.....</b>	<b>S2</b>
<b>2. Supplementary methods.....</b>	<b>S3</b>
<b>3. Supplementary tables and schemes.....</b>	<b>S11</b>
<b>4. Supplementary figures.....</b>	<b>S17</b>
<b>5. Supplementary spectra.....</b>	<b>S30</b>
<b>6. Supplementary references.....</b>	<b>S34</b>

## Supplementary Discussion

### Regarding Scheme 1:

In terms of Scheme 1, the dirigent protein (DIR) and pinoresinol-lariciresinol reductase (PLR) genes are striped red-and-black because homologs of these two genes could possibly be present in *Nicotiana benthamiana*. For *ShDIR*, the top BLAST hit against the predicted *Nicotiana benthamiana* proteome shares 58.55% identity and is predicted to be a dirigent protein. For *ShPLR*, the top BLAST hit against the predicted *Nicotiana benthamiana* proteome shares 75.96% identity and is predicted to be a bifunctional pinoresinol-lariciresinol reductase. Additionally, our previous work<sup>1</sup> showed that *N. benthamiana* indeed had low levels of secoisolariciresinol, the product of PLR, endogenously. We did not, however, see any sign of endogenous matairesinol, the next intermediate in the pathway that is generated by secoisolariciresinol dehydrogenase (SDH).

### Regarding experiments involving candidate laccases:

In addition to investigating the role of the dirigent protein in improving pathway yields in the beginning stages of our work, we also tested candidate laccases (**Table S4**) to determine whether or not the oxidative enzymes already present in *N. benthamiana* are sufficient to accommodate the additional coniferyl alcohol flux. When tested *in planta* alongside the DPT pathway, neither candidate laccase resulted in a significant improvement to DPT yield compared to a GFP control, whether or not coniferyl alcohol was also added exogenously (**Figure S13**). This supports the hypothesis that coniferyl alcohol monomer supply and stereoselective coupling are the key limiting factors, rather than actual monomer oxidation process.

## Supplementary Methods

Note: Parts of this section are adapted from Lau and Sattely 2015<sup>1</sup> as indicated and appear here for clarity.

### Chemicals

Coniferyl alcohol was purchased from Ark Pharm. (−)-Deoxypodophyllotoxin and (−)-yatein, used for standard curves, were isolated from *Anthriscus sylvestris* in previous work.<sup>1</sup> (−)-matairesinol, (−)-pluviatolide, and (−)-burseherin were prepared as described previously.<sup>1</sup> CDCl<sub>3</sub> and C<sub>6</sub>D<sub>6</sub> NMR solvents were purchased from Cambridge Isotope Laboratories.

### RNA-Seq co-expression analysis and coniferyl alcohol biosynthetic gene candidate selection

We utilized the *S. hexandrum* RNA-Seq data and transcriptome generated in previous work<sup>1</sup> to select coniferyl alcohol biosynthetic gene candidates. The transcriptome was first mined for transcripts that had been annotated as laccases or as any of the eight core coniferyl alcohol biosynthetic genes by blastx using the *Arabidopsis thaliana* proteome (The Arabidopsis Information Resource, TAIR) as a database. From this list, up to three transcripts per core gene were selected based on correlation to the DIR transcript expression profile. For the laccases, the top transcript (30060) had a Pearson's r of 0.93 for correlation to DIR expression, while the next two transcripts (1843 and 5587) had values of 0.86 and 0.85. It was determined through analysis of the sequences that these latter two transcripts correspond to the same gene. Both of the two identified candidate laccase genes were ultimately tested (LAC1 = 30060, LAC2 = 1843/5587). For the coniferyl alcohol biosynthetic gene candidates, correlation to the expression profiles of PLR, SDH, CYP719A23, O-methyltransferase (OMT) 3, CYP71CU1, OMT1, and 2-oxoglutarate/Fe(II)-dependent dioxygenase (2-ODD) transcripts was then analyzed for each of the selected transcripts. Of the top three or fewer candidates for each gene, the single transcript whose expression profile most highly correlated, on average, with the expression profiles of the DIR transcript and the other seven gene transcripts was chosen. Notably, the only enzyme for which this selection differed from the transcript that had the highest expression profile correlation with solely DIR was CAD (Table S2). Comparison to CYP71BE54 and CYP82D61 expression profiles was not carried out as these enzymes were not found in the transcriptome generated in previous work, but rather in an online transcriptome dataset from the *S. hexandrum* rhizome. The full-length genes from the selected transcripts were predicted from overlapping transcripts in the transcriptome, and it was determined that each of the eight selected genes' predicted proteins are >70% identical on the amino acid level to the respective characterized coniferyl alcohol biosynthetic genes from *A. thaliana* (Table S3).

### Cloning of coniferyl alcohol biosynthetic genes and laccase candidates from *S. hexandrum* cDNA

Phusion High-Fidelity DNA Polymerase (Thermo Scientific) was used for all PCR amplification steps according to the manufacturer's instructions. All other enzymes used for cloning were purchased from New England BioLabs. Oligonucleotide primers were purchased from Integrated DNA Technologies. Plasmid DNA was isolated from *E. coli* cultures using the QIAprep Spin Miniprep Kit (Qiagen). For a list of primers used for cloning, see Table S5. All gene sequences were amplified from *S. hexandrum* cDNA template generated in previous work.<sup>1</sup> Purified amplicons were inserted into pEAQ-HT<sup>2</sup> (kanamycin resistant) plasmid digested with AgeI and XhoI in an isothermal DNA assembly reaction, as described by Gibson et al.<sup>3</sup> The assembly reaction mixtures were used directly to transform *E. coli* cells, and the isolated plasmids harboring the desired insert were confirmed by Sanger DNA sequencing performed by Elim Biopharmaceuticals, Inc.

For the genes encoding 4CL, HCT, C3H, CCoA-OMT, CCR, CAD, and the two laccases, *E. coli* TOP10 cells (Invitrogen) were used for plasmid isolation prior to transformation into *A. tumefaciens*. DNA excised from agarose gels was purified using the Zymoclean Gel DNA Recovery Kit (Zymo Research).

For the genes encoding PAL and C4H, *E. coli* NEB 5 $\alpha$  cells (New England BioLabs) were used for plasmid isolation prior to transformation into *A. tumefaciens*. DNA excised from agarose gels was purified using the "freeze-squeeze" method.<sup>4</sup>

### **Transient expression in *N. benthamiana* for non-isolation experiments**

pEAQ-HT constructs were transformed into *Agrobacterium tumefaciens* (GV3101) using the freeze-thaw method.<sup>5</sup> Transformants were grown on LB plates containing 50 µg/mL kanamycin and 30 µg/mL gentamicin at 30 °C. Cells were resuspended in 1 mL of LB medium. The suspension was centrifuged at 8000 g for 10 min., and the supernatant was removed. The pellet was resuspended in 10 mM MES buffer, pH = 5.6, 10 mM MgCl<sub>2</sub>, 150 µM acetosyringone, and incubated at room temperature for two hours. Total OD<sub>600</sub> of the *A. tumefaciens* strain mixture suspensions was held at 0.6 (except for the experiment in which various total OD<sub>600</sub> values were tested, the experiment in which different infiltration-to-harvest times were tested, in which the total OD<sub>600</sub> was 3.0, and the experiments for large-scale production and isolation, in which total OD<sub>600</sub> was also 3.0), with each of the individual strains being present at equal concentration unless otherwise noted. *A. tumefaciens* suspensions were infiltrated into the underside of *N. benthamiana* leaves with a needless 1 mL syringe. Plants were grown for five to six weeks under a 16 hour light cycle prior to infiltration. Leaves were harvested five days post-*Agro*-infiltration (except for the experiment in which various infiltration-to-harvest times were tested and the large-scale production/isolation experiments, in which leaves were harvested 7 or 9 days post-agroinfiltration), flash frozen in liquid nitrogen, and stored at -20 °C for later processing. Biological replicates consisted of one leaf from each of three different tobacco plants. Infiltrated leaf areas typically showed some signs of chlorosis (yellowing); leaves expressing GFP as a control also showed a similar phenotype.

For substrate feeding studies, 200 µM of coniferyl alcohol (unless otherwise noted) in 0.1% DMSO in water was infiltrated into the underside of previously agroinfiltrated leaves with a needless 1 mL syringe four days post-agroinfiltration. Substrate infiltration was no more difficult than *A. tumefaciens* suspension infiltration. Leaves were harvested one day later, flash frozen, and stored at -20 °C for later processing.

For all cases in which DIR was not itself tested as a factor in the experiment, a single strain of *A. tumefaciens* harboring the pEAQ-HT-GG:DIR-PLR-SDH construct from earlier work<sup>1</sup> was used. For experiments in which DIR was itself tested as a factor, separate strains harboring expression constructs for DIR, PLR, and SDH were used. In all experiments in which the target product was (–)-deoxypodophyllotoxin or a downstream metabolite, a single strain of *A. tumefaciens* harboring the pEAQ-HT-GG:CYP719A23-OMT3 construct from earlier work<sup>1</sup> was used.

### **Metabolite extraction**

Frozen leaf tissues were lyophilized to dryness (24-48 hours). The samples were homogenized on a ball mill (Retsch MM 400) using 5 mm diameter stainless steel beads, shaking at 25 Hz for 2 min. 20 µL of an 80:20 MeOH/H<sub>2</sub>O solution was added per milligram of dry tissue, and the mixture was heated at 65 °C for 10 min. and then filtered through 0.45 µm PTFE filters before liquid chromatography-mass spectrometry (LC-MS) analysis.

For glycosylated derivatives of (–)-matairesinol and (–)-pluviatolide, an acid hydrolysis protocol was followed as previously described with slight modifications.<sup>6</sup> The methanol extracts were dried under nitrogen, resuspended in an equal volume of 1M HCl aqueous solution, and heated at 95 °C for 2 h. The acid hydrolysis products were extracted into equal volumes of ethyl acetate three times, and the combined organic layer was dried under nitrogen, re-dissolved in 80:20 MeOH/H<sub>2</sub>O, and filtered for LC-MS analysis.

### **LC-MS analysis of metabolite extracts**

Metabolomics samples were analyzed by reversed-phase chromatography on an Agilent 1260 HPLC, using a 5 µm, 2 × 100 mm Gemini NX-C18 column (Phenomenex). Water with 0.1% formic acid (A) and acetonitrile with 0.1% formic acid (B) were used as the mobile phase components at a flow rate of 0.4 mL/min. with the following 41 min. gradient: 0-30 min., 3-50% B; 30-31 min., 50-97% B; 31-36 min., 97% B; 36-37 min., 97-3% B; 37-41 min., 3% B. A coupled Agilent 6520 Accurate-Mass Q-TOF ESI mass spectrometer was used to collect MS data in positive ion mode (parameters: mass range: 100-1700 m/z; drying gas: 300 °C, 1 L/min.; nebulizer: 25 psig; capillary: 3500 V; fragmentor: 150 V; skimmer: 65 V; octupole 1 RF Vpp: 750 V; 1000 ms per spectrum). The first minute of each run was discarded to avoid salt contamination of the MS apparatus. For tandem mass spectrometry (MS/MS) analysis, 5, 10, 20 and 40 V collision energies were used with an m/z window of 1.3 centered on the m/z analyzed. Unless otherwise noted, MS/MS spectra shown are at 10 V collision energy.

Standard curves for (–)-deoxypodophyllotoxin based on UV absorption at 280 nm and ion abundance, and for (–)-yatein and (–)-morelensin based on ion abundance, were constructed and used to quantify amounts produced in *N.*

*benthamiana*. UV absorption at 280 nm and ion abundance were determined by manual integration of UV or extracted ion chromatogram (EIC) traces, respectively, using MassHunter Qualitative Analysis software. For (–)-deoxypodophyllotoxin, the standard curve generated from UV absorption at 280 nm was not utilized for yields less than 50 µg/g of plant dry weight, as UV signal for concentrations lower than this was not well-defined for integration. For yields lower than this, standard curve generated from ion abundance was used, as this curve was linear for yields in this range.

### Statistical analysis of yield results

Remarks about data significance come from the results of Welch's unequal variance t-test. Comparisons resulting in a p-value of <0.05 were interpreted as a sign of a significant difference between the two sets being compared, while comparisons resulting in a higher p-value were interpreted as a sign of no significant difference between the two sets.

### Isolation of (–)-deoxypodophyllotoxin and (–)-morelensin from *N. benthamiana* leaves

Three to five leaves per plant on fifteen to twenty plants were infiltrated with the noted mixture of *A. tumefaciens* strains; harvesting occurred after seven to nine days. After being flash-frozen, leaf tissue was lyophilized for at least 48 hours. Dry leaf tissue was then pulverized into a fine powder with mortar and pestle. 40 µL of tobacco of methanol was added per milligram of dry tissue, and the mixture was refluxed at 65°C for one hour and filtered using vacuum filtration. The methanolic extract was then concentrated *in vacuo*.

(–)-deoxypodophyllotoxin: The dried extract was purified by flash silica chromatography with 9:11 v/v EtOAc/Hxn to yield a yellow residue. After flash purification and crude NMR analysis, samples were further purified by preparative HPLC using an Agilent 1260 Infinity preparative-scale HPLC system with an Agilent 1100 diode array detector and a Clipeus C18 10 µm 250 × 20 mm column (Higgins Analytical), yielding a white powder. Water with 0.1% formic acid (A) and acetonitrile with 0.1% formic acid (B) were used as the mobile phase components at a flow rate of 4.3 mL/min with the following method: 0–10 min, 30–45% B; 10–39 min, 45% B; 39–40 min, 45–97% B; 40–55 min, 97% B. TLC (EtOAc/Hxn, 9:11 v/v):  $R_f = 0.3$ .  $^1\text{H}$  NMR (600 MHz, Benzene- $d_6$ ) δ 6.61 (s, 2H), 6.48 (s, 1H), 6.42 (s, 1H), 5.32 (d,  $J = 1.4$  Hz, 1H), 5.29 (d,  $J = 1.4$  Hz, 1H), 4.42 (d,  $J = 4.9$  Hz, 1H), 3.79 (s, 3H), 3.63 (dd,  $J = 8.1$  Hz, 7.4 Hz, 1H), 3.51 (s, 6H), 3.04 (dd,  $J = 10.6$  Hz, 8.3 Hz, 1H), 2.35 (m, 1H), 2.18 (dd,  $J = 16.0$  Hz, 5.4 Hz, 1H), 1.95 (dd,  $J = 13.9$  Hz, 4.9 Hz, 1H), 1.86 (dd,  $J = 15.9$  Hz, 11.7 Hz, 1H).  $^{13}\text{C}$  NMR (151 MHz, Benzene- $d_6$ ) δ 174.2, 153.7, 147.4, 147.2, 139.1, 136.7, 131.4, 129.0, 110.9, 109.8, 108.6, 101.2, 71.3, 60.4, 56.3, 47.3, 44.2, 32.8, 32.6. This compound has been isolated previously from the dried steam bark of *Bursera fagaroides* var. *fagaroides* (7.9 µg/g DW isolated yield)<sup>7</sup> and from whole plants of *Anthriscus sylvestris* (no isolation yield reported).<sup>8</sup>

(–)-morelensin: The dried extract was purified by flash silica chromatography with 30% ethyl acetate in hexanes and further purified by preparatory HPLC using the method described above.  $^1\text{H}$  NMR (500 MHz, Benzene- $d_6$ ) δ 7.19 (s, 1H), 6.51 (s, 2H), 6.47 (s, 1H), 6.44 (s, 1H), 5.32 (d,  $J = 1.3$  Hz, 1H), 5.29 (d,  $J = 1.3$  Hz, 1H), 4.42 (d,  $J = 4.9$  Hz, 1H), 3.66 – 3.58 (m, 4H), 3.36 (s, 3H), 3.03 (dd,  $J = 10.7$ , 8.4 Hz, 1H), 2.33 (m, 1H), 2.16 (dd,  $J = 15.9$ , 5.3 Hz, 1H), 1.96 (dd,  $J = 13.8$ , 4.9 Hz, 1H), 1.86 (dd,  $J = 15.9$ , 11.7 Hz, 1H).  $^{13}\text{C}$  NMR (151 MHz, Benzene- $d_6$ ) δ 174.2, 149.8, 149.4, 147.4, 147.2, 134.1, 131.9, 128.9, 123.2, 116.0, 111.8, 110.9, 108.6, 101.1, 71.4, 55.9, 55.6, 47.3, 43.7, 32.8, 32.5. This compound has been isolated previously from the dried exudates of *Bursera morelensis* (from 680 g of dried exudates, no isolation yield reported)<sup>9</sup> and the fruits of *Anthriscus sylvestris* (10 µg/g DW isolated yield).<sup>10</sup>

All purification procedures were carried out with at least reagent-grade solvents. Standard column chromatography technique with SilicaFlash® P60 silica gel (40–63 µm) was used for purification.

### NMR Analysis

All spectra were analyzed with MestReNova NMR analysis software.

$^1\text{H}$  NMR spectra were recorded at room temperature on a Varian 400 (400 MHz), an Inova 500 (500 MHz), or an Inova 600 (600 MHz) spectrometer. The Varian 400 was only used for NMRs taken prior to preparatory HPLC. Chemical shifts are reported in ppm downfield from tetramethylsilane using solvent resonance as an internal standard (Benzene- $d_6$ : 7.16 ppm).

Proton-decoupled  $^{13}\text{C}$  NMR spectra were recorded at room temperature on an Inova 600 (151 MHz) spectrometer. Chemical shifts are reported in ppm downfield from tetramethylsilane using solvent resonance as an internal standard (Benzene- $d_6$ : 128.1 ppm).

## *S. hexandrum* Gene DNA Sequences

### Laccase 1 (LAC1)

ATGGGTTCACTTCCAAGGGGAGCATTCTCATTGCTTTACCTGTTATGGTCTGCCAGGGCTTATCA  
ATGCAAAGAATGCAGAAACTACAAGGCCTACAAGTGTGATATCAAATTGCAGGATGTGACTAGATTA  
TGTCAATTCCAAGAGTATCGTAACTGTCAATGGACAATTCCAGGTCTCAAATAATAGCGAGAGAAGG  
AGACCGAGTGTATCGAGGTGGTTAACCATGTCCAGGACAATATCACCTTGCCTGGCATGGAATT  
GTCAATTTCGAAGTGGATGGCTGATGGACCTGAATATATAACCCAATGCCCATACAGACAGGCCGA  
AGCTATGTCTACGACTTACCCATTGTTGAGCAGAGAGGCACTCTGGTGGCATGCCACGCTACATG  
GCTGAGGTCTACTGTTCATGGAGCCATCATTATCCTCCCTAAGAACGGTGCCAATATCCTTCTCAA  
ACCTTACAAGGAAATGCCTATTATTCGGAGAGTGGTCAATTCAAGATACCGAAGCGATTATGAGAG  
AGGCTGACAAACGGGAGGAGTCCTAATGTCTGAAAGCTTACACTATTAAATGGACTACCAGGACCT  
TTGTACGACTGTTCTCTAAAGATACTCAAGCTGAATGTAAGCCTGGGAAGACATATCTTCTCCGA  
ATAATAAAATGCTGCACTCAACGATGAACCTTCTCAGCATAGCCGCCACACTCTTACCGTTGCGAA  
GCTGATGCAGTCTATGTGAAGCCATTGAAACTGATACCACCTTATCAGCCCAGGCCAAACTACAAA  
TGTCTTCTCAAGACCAGACCAACTACCCCCAAGCCACCTTCTTATGGCTGCTAGACCTTATGCAAC  
TGGCCTTGGCACCTTGACAACACTCCACCTGCTGGATTCTGAATACGAAAACCTTCAAGGCAAGA  
CTAAGAAATTCATTGCTAACACCAACCCATTCCCCATCTAAATGATATCCCATTGCTACAAAGTTT  
CCAAAAGTCATCGTAGCCTAGCCACTTCTGAGTCCGGTAATGTCCTCAAGATTGTCATAAACGCT  
TCTATTTCACGGTGGGATTGGATCAAATCCTTGTCCACCCACCATACATGTCAGGGCCTTATGGAG  
CCAAGTTGCAAGCTCAATAATATCATTGACTTCCAACGACAGCTACTTCAAACCCATT  
TCTTGGAAAATCCATTGGTGTACTCTGCCGATTTCCCTACTAAACCACGTATCCCTTGCACTACA  
CTGGAAACCCCTCAAACAATACTAATGTGGCAGTGGCTGTAATTGGTCAATTGTCATGTTACCATTCAACACA  
AATGTGGAGCTAGTGTGCAAGGTACTAGCATTGGTGTGAGAGGCCATCCTCTGCACCTCATGGA  
TATAACTTTATGTTAGGTCAAGGTTGGTAACTATGATCCCATTGAGGATCCTAGTAACCTCAAT  
CTTGTGGACCCATTGGAGAGAAACACAGTGGTGTGCCGCTGGTGGTTGGCTGCATTGATTTCTC  
GCAGACAACTCTGGTGTATGGTCAATTGGAAGCCCACACAACCTGGGCTTGGCCAT  
GGCTGGATAGTCATGGATGGAGCTCCCAATCAGAAATTGCCCTCACCGTCTGATCTCCCCA  
GTGCTAA

### Laccase 2 (LAC2)

ATGGCATCTCTTGTITCCCTCAAGGGGAGCATTCTCATTGTTTACATGTTATGGTCTGCTGG  
GGCTTATCAATGCAAAGCATGCAAGAAACTACTAGATATTACAAGTTGATATTAAATTGCAGGCTGTG  
ACTAGATTATGTCAACCCAAGAGTATCGTACTGTCAATGGAGAATTCCAGGTCTCAAGTAATAGC  
GAGAGAAGGAGACCGAGTGTGATTGTTGAGGTGGTAACCATGTCGAGGAAATGTCACCTGACTGG  
CATGGGATTCGTAACCTCGAAGTGGATGGCTGATGGACCTGCATATGTAACCTCAATGCCCAACT  
TACAGGCCAAAGCTATGTATACAACCTCACCATTGTAAGGCAGAGAGGCACCTCTGGTGGCATGCC  
ACACTTCATGGCTAAGGGCGACTCTTACGGACCGATCATTATCCTACCTAACGCCGATAGTCATATC  
CTTCCCCAGACCTTACAAGGAAATACCGTTATTCTCGAGAGTGGTTAATGCGGATACCGAAGCT  
GTTATCAAGGAGGCTCTACAAACAGGAGGTGGCTTAATGTCATGCTTACACCATTAAATGGACT  
CCCAGGGCCCTGTATACCTGTTCTAAAGATATATTCAAGTTGAGTGTGAAGCCTGGCAGACAT  
ATCTTCTCCGAAATAATAATGTCACTCAATGATGAGCTTCTTCAGCATAGCCGACCAACTCTTA  
CCGTTGTCGAAGCTGATGCACTGTTGAAACTGATACCACCTTATCAGCCAGGCC  
AAACCACAAATGTTCTCAAGACCCAGACCATACTACCCAAAGCCACCTTCTTATGGCTGCTAGA  
CCTTATGCAACTGGCTTGGTACCTTGTACAACCTCCACCACTGCTGGAATTCTCGAATACGAAAACCTT  
TCAGACAAAGACTAAGAAACTCCATTGCTAACACCAACCCATTCTAAATGATATTCCATTGCT  
ACAAAATTTCAGAGTCATCGTAGCCTAGCCACTCTGAGTTCCGGTTAATGTCCTAACGATTGTC  
GATAATCGCTTATTTTACCGTGGGATTGGGATCAAATCCTGTCCACCCAAAGATACTGTCAGGG  
CCTTATGGAGCCAAGTTGCAAGCTGCACTGAGTAATAATATCATTGTAACCTCCAACACTACGCTACTT  
CAAACCCATTCTTGGAAAATCACACGGTGTGACTCCGCTGACTTCCCTACTAAACCCACCTATCCCT  
TTCGACTACACTGGAAACCCACCAAAACAATACTAATGTGGCAGTGGCACTAATGTGGTGTGATGTTACC  
ATTCAACACAAGTGTGGAGCTAGTGTGCAAGATACCGACATTCTGGTGTGAGAGGCCATCCTCTC  
ACCTTCATGGATTCAATTCTTATTGTAAGGTCAAGGTTGGTAACATGATCCCAGCAAGGATCCTG  
TTAACTTCATCTTGTGGACCCATTGGAGAGAAACACATGTGGTGTGCCCTGCTGGTGGTTGGGTTGCCA  
TTCGATTCTCGCAGACAATCCTGGAGTATGGTCAATTGCACTGCCATTGGAAGCCCACACAAGCTGG  
GGATTGAGGATGGCTGGATAGTTATGGATGGAGCTCCCAATCAGAAATTGCCGCCCTCACCATC  
TGATCTTCCCAAGTGTAA

PAL

ATGGGAAGTTAGTGCAGAAAACCGAATTCAATTCCCAGCTTGCATAAAAGAACCCATTGAA  
TTGGGGTATGGCAGCAGATGCATTGAAAGGAAGTCACCTGTAGGAGGTGAAGAAAGATGGTCAAGGAG  
TATAGACAAGGGGTGTGAGATTAGGAGGGGAGACCCTACTATATCCCAAGTAGCTGCAGTTGCTTC  
CCATGATCAAGGTGTCAAGGTGGAACACTCAATGAGTCAGCAAGGGCAGGTGTTAAAGCTAGTAGTGATT  
GGGTTATGGAGAGCATGAACAAAGGTACAGATAGTTATGGTGTACAACACTGGATTGCTACTTCT  
CATAGAAGGACAAAGCAAGGTGGTCACTGCAGAAGGGAGCTCATTAGATTTGAATGCCGAGTCTT  
TGGCAATGGAACAGAACATCCACCCACACGTTGCCACTCAGCAACAAAGGGCAGCAATGCTAGTCAGA  
ATCAACACTCTCCTCCAAGGCTACTCAGGCATTGTTGAAATCTGGAAGCCTGACCAAATTCTC  
GACCACAAACATTACCCATGCTTACCTCTAAGGGTACAATCACTGCATCCGGTGTACTTGCTCCCTCTA  
TCCCTACATTGCTGGTCTATTGACCGTAGGCTTAATCGAAAGCAGTGGGACCTAAAGGTGAAACCCCT  
CAATGCAGAAGAAGCATTACCCCTAGCTGGTATTAAATGGTGGCTTGTAACTGCAACCCAAGGAAG  
GACTTGCATTAGTCATGGTACTGCAGTGGGCTGGACTTGCTCAATGGTCTATTGAGGCAAACA  
TACTAGCTGTTCTATCTGAAGTTATTCAGCAATTTCGAGAAGTAATGCAAGGGAAACCAGAAATT  
CAGATCATTTGACACATAAGTGAAACATCACCAGGCCAATCGAAGCTGCAGCAATCATGGAACAT  
ATTCTTGATGGTAGITCATATGTAAGGAAGCAAAAAGCTACATGAAATGGATCCTCTACAGAACCC  
AAAGCAAGATAGATACGCTCTACGAACCTCTCCACAATGGCTGGACCTCTGATGAAAGTAATTGAT  
CATCTACCAAGTCATTGAGCGTAAATCAATTGCGTCAACGATAACCCATTGATGATGCTCCAGA  
AACAAAGGCTCTCCATGGTGGAAATTCCAAGGGACGCCAATTGGAGTTCAATGGACAACACTAGGTT  
GGCTATTGATCGTAGGAAAGCTTGTGTCACAATTCTGAGCTTGTAAATGACTTCTACAACAA  
TGGGCTGCCCTCAAACCTTATCTGGTGGAGGAATCCTAGCTGGATTATGGACTCAAGGGTGTGAGA  
TTGCTATGGCTTCCACTGTTCTGAACCTCAATTCTAGCAAACCCAGTGACTAATCATGTTCAAAGCG  
CAGAACACACACAAAGATGCAACTCTTGGTTAATTCTCGAGAAAAACGCCAAGCAGTA  
GATATCTTGAGGCTCATGCGTCTACTTCTAGTTGGCTTGCAGCAATTGACTTGAGGCATATA  
GAGGAGAATATGAAGAGCTGTTGAAGAACACAGTTAGTCAGTGGCTAAGAAAGTACTAACAAATGG  
GTGTGAACGGGGAGCTTCACCCCTCAAGATTGAAACTTGCTCAAGGTGGTGCAGAGAGAA  
TACACTTCCCTACATTGACGATCCATGCCCTCGCTACTTACCCCTCTAATGCAAAACTCAGAGAAC  
CTCGTGGAACATGCGTTGCTAAATGGAGAGAGCGAGAAAGATCCAAACACTTCAATCTCCAGAAGAT  
CAATACATTGAGGAGGAACCTCAAGGCCCTGTTGCCAAAGAGGTTGAGAGTATAAGAAGTGTATTG  
AGAGTGGAAACCCAGCAATCCAAACAGGATTAGGAATGTAGGTACACCCACTGTACAAGTTGTA  
AGGGAGAATTGGAACTGAGCTGTTGACTGGCGAGAAGGTCCGATACCAGGAGAGGAGATTGACA  
AGGTGTTTCAGCGATGTGCCAAGGGAAAGTTGATTGATCCACTTCTAGAATGTGTTGCAGAGTGGAAAC  
GGTAGCCCTTCTATCTGTTAG

C4H

ATGGATCTCCTCATCTTAGAAAAAACTCTCATCGGCCTATTGTCGCCATCATGGCGCTATCGTAATC  
TCTAAACTACGCGGTAAAGCGTTCAAACACTCCCTCCAGGACCCATTCCGTTCCCATTTCGTAACCTGG  
TTACAAGTCGGAGATGACTTAAACCACCGAAACTTAACCGATTAGCTGTAATTCCGGCAAATCCT  
CCTCCCTCCGATGGCCAACGTAATCTCATGCGTTCTTCCCCGACTTGGCCAAGAGGTTTACA  
TACTCAAGGCCTGAGTCGGTCCCAGCCGAACGTTGTTGACATCTCACCCGCAAAGGAC  
AAGACATGGTTTCACCGTGTATGGCAACACTGGAGAAAAATGCGGAGAACATGACCGTTCTTT  
TTCACTAATAAGGTTGTTCAGCAGTATAGATTGGATGGAGGATGAGATCTCTAGAGTTGAGGA  
TGTGAAGAAGATGCGTGAAGCCGCGACGAATGGGATTGTTGAGAACAAAGGTTGAGCTTATGATGT  
ATAATAATATGTATCGGATTATGTTGATAGGAGGTTGAGAGTGTGATGAGCTGTTGAGGTT  
TGAAGAAATTGAATGGGGAGAGGAGTAGGTTGCTCAAAGCTTGTATTATAATTGTTGATT  
CCTATTGAGGCCTCTGTTGAGGAAGTATTGAAAGATGTGAGAGATATGCAGGAGAACGGTTGAA  
GCTGTTCAAGGATTACTTCTGATGAAAGGAAGAAACTAACGAGGCACAAAGGTTGAGAACGTTGG  
CTAAAATGTGCCATAGATCACATTGTTGATGCTCAAGAGAACGGAGAGATTATGAAGATAATGTTCT  
TTACATTGTTGAAACATCAATGTTGAGCAATTGAGACTACACTGTGGTCGATTGAGTGGGAATCG  
CAGAGCTGTAACCCACCCGAAATCCAACAAAAACTTCGTCATGAGCTGACACAGTACTTGGACCA  
GGCGTGCACATCACCAGCCAGACACCCACAAGCTCCCATACCTCAAGCTGTTATCAAAGAGACCC  
CCGACTCCGATGGCCATCCCATTGCTAGTCCCACACATGAACCTCAGGATGCTAAACACTCGGTGGCTT  
TGATATCCCAGCTGAAAGCAAAATCTGGTTAATGCTGTTCTGCAAACAAACACCCGAAACACTGGA  
AGAATCCTGAAGAATTCCGTCGGAAAGGTTCTGGAAGAGGAAGCTAATGTGGAAGCCAATGGTAA  
TGACTTCAGGTACCTGCCATTGGTGTGGTAGGAGGAGTTGCTTGGATTATACTCGCATTGCCAAT

TCTGGGAATTACTTGGGCCGTTGGTCAAAATTGAGCTTTGCCACCCCTGGCAAGACAAGCT  
TGATACTACTGAGAAAGGTGGCAGTTAGCATGTTATTCTGAAGCATTCCACCATTGTTAAGCC  
CAGAGTGTGTTGA

4CL

ATGGAGACTCCTCCCCAAGAGTACATATTCAAGGTCAAACCTCCTGATATCTACATCCCTACTCATCTC  
CCTCTACATTCCATTGCTTCAGAAAACATTCTAAGTCAGTGACCGGCCATGATCAATGGTGC  
ACTGGAGAAATCCATACTTACGCCAGTCGAGCTCATTGCGTAAAGTCGCTTCCGGACTGAAAAAA  
ACTGGGCATCAACCAATACGATGTTATCATGTTACTACTTCCAATTGCCCGAATTGCGTCTCACATT  
TCTCGGTGCATCGTACCTCGGGCAATGACAACAGCAGCGAATCCATTCTACACGCCGCTGTAACATT  
CCAAACAGCGAAAGCGTCCAACGCGAAGCTGTGTTAACACAAGCTTCCACGTGAGCAAACACTCACC  
GAGTTGACGTGAGAATGATATTAAAGATTATGTGATTCTCCTCCAGCAGAGGAGAATTGGTT  
ACATTCTGAGTTAACTCACTCGGACGAGACTGAATTAAACTCAGGTGGAGATTGCCCGGACGATG  
TTGTGGCACTGCCATTATCATCCGGACGACTGGGTTACAAAAGGTGTTATGCTGACGCATAAAGGG  
TTGGTGAAGTGTGGCACAGCAGGTTGATGGGAGAATCCTAATTGATACATGCACAGTGGAGATGT  
GATGATATGTGTTGCTATGTTACATTACTCGCTTAACCGGTTGTCGGGTTACGAGTT  
GGCGCGCGATTGATTATGAGTTGATATTACAGTTGAGTTGAGCTTCAAGGTGAGGTTGAGTATAAG  
GTGACGATTGCGCCGTTGTTCTCGATCGTGCCTGCCATTGTCAAAAACCCAGCTTCTGAAAAAATAC  
GATCTGCGTCGATTGGACGGTCATGTCGGTGCACCTATGGGAAGGAGCTGAGGACAATT  
GAGGGCTAACGTTCTAACGCCAATTGGACAGGGATATGAAATGACAGAACAGCAGGGCCGTTG  
GCAATGTGTTGGCTTGCAAGGAGCCATATGAAATTAAATCAGGTCATGTTGAACTGTTGTTAG  
AAATGCTGAGATGAAGATTGTGGATCCAGATACCGGTTATCCCTGCCTCGAAATCAACCCGGAGAGA  
TTGTATTAGGGGAGCTCAGATCATGAAAGGTTATGTTAATGATCCTGAGGCAACTGCTGCAACCATA  
GACAAAGAAGGTTGGTACATACAGGTGACATAGGCTATATCGATGACGACAATGAGCTTTTATCGT  
GGATCGATTGAAGGAGCTGATTAAGTACAACAGGGTTCCAAGTTGACCTGCTGAACCTCGAAGCCATGC  
TGATTAACCACCCAAACATCTGATGCTGTTCTCAATGAAGGATGATGCTGCAAGGAGATT  
CCCGTGGCTTCATCGTGAAGATCAAACGGTCACAGATTACTGAGGAGCAGAAATCAAGCAATATGTCG  
GAAACAGGTGGTGTCTACAAAAGGATAGGCAAAAGTGTCTCGTCAACACATTCCGAAGGCCAT  
CGGGCAAAATCCTGAGGAAGGATTAAAGAGCAAGGCTAGCAGCTGGTCTCCTAATTAG

HCT

ATGATTATTAACGTGAAGGAATCAACAAATGGTACCCCCAGCAGAAGAAACCCAAAGGAGAAGCCTCT  
GGAACCTAACTGAGACTTGGTCGTGCCAGAATGCACACACCTAGTGTAACTTCTATAGACCCAAATG  
GTTCCCTCTAGCTTTTGATTCCAAGGTGCTCAAGGAGGCCTTGAGTAAGGCCCTGTACCGTTCTACC  
CTATGGCTCAAGGCTAACAGAGGGATGAAGATGGTAGGATTGAGATTGATGGACAAGGGGT  
GTTACTTGTGGTGGCGGAAACCGGATCCGTTATCGATGATTGCGGGATTGCTGACCTACTTGGAGCT  
CCCGCAACTCATTCCGGCGTTGATTACTCCGGGACATTGTCATATCCACTTCTGTTGCAAGGT  
GACTCATTCAAATGTGGCGAGTCTCACTGGAGTTGGATGCAACACCAACGTCGCGGATGGAGCCT  
CTGGCCTTCATTCATCAACACATGGTCTGACATGGCACGTGGCCTGATCTCACCATCCCACCCCTCA  
TCGACCGCACCCCTCTGCGTGCACGTGACCCACCTACACCTGTTAACACGTCGAATACCAACCTC  
CTCCCTCTATGAAAACCCAAATTCTCAATGACACCCACTGAAACATCGTAGCTATTCAAGATGA  
CTCGTGAACAAGTAAGCACCTCAAAGCCAAATCAAAGAGATGGCAATACCATAAATTACAGTTCT  
TATGAAATGTGGCTGGACATGTATGGCGCTCCGATGTAAGGCGGTGGTTATCTGATGATCAAGA  
AACCAAAATGTATATTGCTACTGATGGACGATATAGATTGCGCCAGCTCTCCACCAAGGTTATTGG  
CAATGTGATCTTCACAACCACCCAACTTCAGTATCCGGGACTTGAACACAAAGCCCTACACTATG  
CCCGAGGTGAATTACAATGCACTAGTGCAGTACGGACGATGAGTACATGCGATCAGCCCTCGATTAC  
CTAGAAATGCACTGATCTTCAGCCCTGTGCGTGGGGCACACACATTAAATGTCCAATCTGGC  
ATCACAAGCTGGACGCCCTCCGATTGATGCTGATTGTTGGTGGGGCACCCATTGATGGGA  
CCAGGAGGGATTGCTTACGAGGGACTCGCATTGTTGCAAGCCCTAACGGTATGGCAGCTGTC  
CTTGGCTATTCTGTTACAGCCGAGCACATGAAACATTGAAAAAATTGATGAGTTTAG

C3H:

ATGGCTCTCTCGTGTCTCTCATCCCCATTTCATCATCCTGCTTACAACATCTACCATCACCTGA  
GATTCAAACCTCCACCAGGCCAAAACCATGGCCCATAGTAGGAAACTTGTACGACATCAAACCAAGTG  
AGATTCCGGTGTGTTGCAAGATGGCTCAGATATACGGCCCCATCTGAGTCTGGTTGGATCAACG  
TTGAACGTTGTGTTGACAATTGCAAGGAGGTGTTGAAGGAGAAAGATCAACAGTTGG

CTGATAGACATAGGAGCAGATCAGCTGCCAAGTTAGTAGAGATGGACTGATTGATTGGCTGAT  
TATGGGCCTCACTATGTTAAGGTGAGGAAGGTTGACGCTTGAGCTTCTCCTAAAAGGCTGAA  
GGTCTTAGGCCATTAGAGAACATGAAAGTTACTGCCATGGTGAATTCATCTTCATTGATTGACCAA  
CCAGAAAATTATGGCAAGAGTGTGGTGAAGAACAGTACTGGGTGAGTGGCATTCAACAATATAA  
CACGACTCGCCTTGGAAAACGCTTGTGAATGCAGACGGCGTATGGATGAGCAAGGAGTTGAATT  
AAGGCAATTGTCGAATGGATTAAAGCTTGGTCATCACTGCCATGGCTAACACATTCCATGGCT  
CCGCTGGATGTTCCCATTAGAGGAGGGGCATTGCTAACATCAGGCTCGTAGGGATCGTCACTA  
GAGCTATTATGGAAGAGCACACTGCTGCCGTAAAAGAGTGGTGGTCCAAGGAACATTGTCGAT  
GCATTGCTCACTCTCAGGAGAAATATGATCTCACTGAGGATACCATCATGGGGCTCTGTGGGACAT  
GATCACTGCTGGATGGACACAACGCAATATCAGTTGAATGGCTATGGCGAGCTAACAGAACATC  
CAAGGGTCCAAGAAAAGGCCAAGAAGAACATTGGATCGAGTCATTGGTTGAACGTGTCATGACTGA  
GACCGACATCTCAACCTACCTACAATGTGTCGAAAGAGGCACCTCCGGTTGCACCCTCCAA  
CTCCACTCATGCTCCCCACAAGGCCAATGCCAATGTTAAGATAGGTGGCTATGATATCCCCAAGGGA  
TCCAACATGCATGTCATGTATGGCTGTGGCTCGTATCCTGCTGTATGGAAGGAACCACTAGAATT  
TAGGCCAGAACGGTCATGGAAGAGGATGTCGATATGAAGGGACATGATTACAGATTACTACCATTG  
GTGCTGGCGACGAATCTGCTCTGGGCACAACGGTATTAACTGGTGCATCAATGATAGGTAC  
TTGTTGCACCATTTCGTTGGCTCCACCTGAAGGGTTGGCCGAAGAGATCGACATGTCAGAAAA  
CCCTGGAATGGTAACATACATGACGACCCATTGCAAGCAGTCCTACACCTAGGTTGCCCTCACAGC  
TATACAAACGCATGCCGTGGACATGTGA

#### CCoA-OMT

ATGGCAACCAACCAGACTGAGCAAACAAACAGCCAGAACATGAGACATGCTGAGGTTGGCACAAGA  
GTCTTCTCCAAGCGATGCTCTTACCAAGTACATCGAAACAAGTGTATCCAAGGGAGCCACAA  
GCCATGAAAGAACTCCGTGAGATTACAGCCAACCATCCTGGAACCTGATGACAACCTCTGCTGATGA  
GGGACAGTTTGGCGATACTTCTTAAGTTGATTAACGCCAAAACACCATGGAGATTGGTTTTTAC  
TGGATATTCTCTCCTGCAACTGCTCTGCTCTGAAGATGGAAAGATCTGGCGATGGACATCAA  
CCGGGAGAAACTACGAAATTGGTCTGCCGTGAATTCAAAAAGCTGGAGTGGCTCATAAGATTGATT  
GAGAAGGACCAGTTGCCGTTGACCAAATGATCGAAGATGAGAAGTACCATGGATCGTCTGAT  
TTCATATTGTCGGACGCTGACAAGGATAACTATCTCAATTACCAAGAGATTGATGATTGGTGA  
GGTTGGAGGAGTGTAGGATATGACAACACTCTGAGTGGAGCTCAACAGGATTGGCAGCAGA  
CCATGCGCAAGTACGTGAGGTACTACAGGACTTCGTATGGAGCTCAACAGGATTGGCAGCAGA  
TCCAAGGATCGAGATTGCCAGCTCCAGTCGGCGATGGAATCACCCTGCGCTCGTCACTG

#### CCR

ATGCCTGTTGACAGTTGTTACCTCCAGTCAAACTATCTCGTCACCGGCCGGAGGTTCATCGCT  
TCATGGATGGTCAAGCTTGTGGAGAGAGGTTAACTGTTAAAGGAACATGTCAGAAATCCAGATGA  
TGATGCAAAACGCTCATTGAAGGCTTGAAGGAGCGAATGAGAGATTGATTGATTGAAAGCGG  
ATCTCCTGACAGGGAGAGCCTCATGAAGCGATTCAAGGCTGATGGAGTTCCACACTGCATCA  
CCTGTTACTGATGATCCGAACAAATGGTGGAGGCCAGCAGTGAACGGAACGAAATATGTGATTGATGC  
AGCGGCAGCTGGTGTGTCGAGTTGTTACCTCATCAATTGGTGCCTGACCATGGATCCTAA  
TCGGGGTCTGATGTTGTTGATGAGAGTTGAGTGGAGCGATCTGAATTCTGCAAAAGCCACAAAGA  
ATTGGTATTGTTATGGAAGGCAGTGGCTGAGAAAGCAGCACTAAAGATGGCAAAAGAGAAAGGAGT  
GGACTTGGTGGTGAACCCAGTACTGGTGTGGTCCACTCCTCAACCGACTGTTACGCTAGCA  
TTGTTCACATACTCAAGTATCTCACTGCCCAAGACCTACGCCAACTCTATTGAGCTTATGTGG  
ACGTGAGGGATGTGGCATTAGCTCACATCCTGTTACGAGACTCCCTCTGCGCAAGGGAGGTATCTG  
TGTGCTGAGAGCGTGTGCTCACCGTGGAGATGTGGTGAATACTATCCAAGTTCTCCAGAATACCCT  
GTGCCAATGAAGTGTGGATGAAGTGAAGCCAAGAGTTAAACCATAAGTTCAAACCAGAGG  
TCAGGGACATAGGCCCTGGAGTTACACCAGTGAAGCAAAGTTGTACGAGACCGTGAGGAGCCTACA  
GGAGAAAGGCCATCTCCATCCCTACCCAAACGACGAATCCCTCCGCATTCACTGTTGA

#### CAD

ATGGGTAGTATTGAAGTTGATAGAACCAACTACAGGTTGGCTGCTAGAGACCCATCTGGATTCTGTC  
TCCTTACACTTACCCATTAGGAGCACCGGCCGAAGATGTGTACATCAAGGTCTCTGTGGTGT  
TTGCCACACAGATATTCAACCAAGCTAAAATGATCTGGCATGTCAAACACTACCCATGGTCTGGC  
ACGAAGTTGTGGTGAAGTACTAGAGGTTGGATCAGATGTGACAGCATTCACTGAGACAGAACAGT  
GGGAGTAGGATGCATCGTTGGATCTGCAAGCAGTGTCTTCACTGAAATCAAAGATCGAGCAGTATT

GTAACAAGAGGATTGGTCTTATAATGATGTTACGTTGATGGAAAGCCTACTCAAGGAGGCTTGCT  
GGTGCATGGTGTGATCAAAGTTGTGGTGAAGAATACCCCAAGGACTGGAGCCAGAACAGCAG  
CACCACTACTATGCGCAGGGCTGACAGTGTACAGCCCATTGAGGCATTACAATCTGACCAACAGTGG  
CTTAGAGGAGGCCATACTGGGCTTGGAGGAGTGGTCATATGGGTGTGAAGATAGCAAAGGCCATGG  
GCCATCATGTAACAGTCATAAGCTCTTCTGAAAAGAAGAAAGAGGAAGCTATGGCACATTGGGTGCT  
GATGCTTACCTCGTGANCTGAGCTCAGATGCTGATCAGATGCAGCAGGCTGCGGATAGCCTGATTACACT  
GGACACTGTGCCTGTTGCCATCCACTTGAGCCCTATCTTCTTGTGAAGGTGGATGGAGGTTGAT  
CATCATGGGTGTCATTGCTCAGCCTTGAGTTCTCTATGCTCATGCTGGGGAGGAAGTCGAT  
TTCTGGAACATTCAATTGGAAGCATAGAGGAGACCGGAGAGATGCTCAAGTTGCAAGGAGAAAGGG  
ATAAGGTGATGATCGAAGTGGTAAGATGGACTATATCAACAAAGCTTGAAAGGTTGGAGAAGA  
ATGATGTGAGATACAGATTGTTGATGTTGCTGGTCTTATAATGATGTTACGTTGATGGAAAGCC  
TACTCAAGGAGGCTTGCTGGGCCATGGTTGATCAAAGTTGTGGTGA

## Supplementary Tables and Schemes

**Table S1.** Selection of putative coniferyl alcohol biosynthetic gene candidates according to co-expression with etoposide aglycone pathway biosynthetic enzymes. Brighter yellow colors indicate higher expression profile correlation to that of the dirigent protein gene, while darker colors indicate lower expression profile correlation. Numbers in boxes are Pearson's r values calculated from linear regressions of expression levels (effective counts, TMM-normalized, log<sub>2</sub>-scaled, and median-centered) for the transcripts being compared measured at four time points post-leaf wounding across three biological replicates (See Supplementary Reference 1 for original RNA-Seq data). The averages on the right are averages of the Pearson's r values for their respective columns. Selected transcripts are marked with an underline. Some transcripts may be from the same candidate gene.

Transcript	Candidate Name	Correlation Coefficients (r)								Average
		DfR	PLR	SDH	CYP779A23	OMT3	CYP14C1	OMT1	2-ODO	
<u>31350</u>	<u>PAL_1</u>	0.850	0.524	0.949	0.823	0.920	0.867	0.935	0.719	<u>0.823</u>
11710	PAL_2	0.814	0.443	0.939	0.798	0.897	0.813	0.940	0.665	0.789
3663	PAL_3	0.813	0.399	0.945	0.783	0.915	0.786	0.957	0.655	0.782
<u>3381</u>	<u>C4H_1</u>	0.349	0.222	0.624	0.508	0.461	0.246	0.584	0.259	<u>0.383</u>
3197	C4H_2	0.267		0.651	0.579	0.420	0.322	0.573	0.179	0.371
<u>6619</u>	<u>4CL_1</u>	0.805	0.351	0.983	0.875	0.906	0.790	0.966	0.707	<u>0.798</u>
6618	4CL_2	0.789	0.334	0.966	0.845	0.906	0.797	0.974	0.653	0.783
33034	4CL_3	0.460	0.280	0.867	0.854	0.614	0.692	0.799	0.303	0.609
<u>32941</u>	<u>HCT_1</u>	0.785	0.445	0.955	0.825	0.885	0.783	0.970	0.605	<u>0.782</u>
31386	HCT_2	0.574	0.463	0.916	0.923	0.666	0.737	0.834	0.437	0.694
<u>32010</u>	<u>C3H_1</u>	0.732	0.478	0.945	0.844	0.841	0.814	0.948	0.537	<u>0.767</u>
30934	C3H_2	0.577	0.571	0.930	1.000	0.671	0.898	0.827	0.445	0.740
31211	C3H_3	0.458	0.712	0.103	0.085	0.298	0.389	0.137	0.420	0.325
<u>5680</u>	<u>CCoA-OMT_1</u>	0.873	0.455	0.894	0.721	0.937	0.747	0.962	0.697	<u>0.786</u>
5679	CCoA-OMT_2	0.858	0.480	0.836	0.643	0.908	0.728	0.913	0.687	0.757
3150	CCoA-OMT_3		0.504				0.161			0.071
<u>13434</u>	<u>CCR_1</u>	0.826	0.490	0.903	0.832	0.860	0.764	0.881	0.773	<u>0.791</u>
294	CCR_2	0.723	0.467	0.986	0.921	0.840	0.863	0.957	0.556	0.789
293	CCR_3	0.702	0.450	0.972	0.912	0.825	0.840	0.947	0.538	0.773
12724	CAD_1	0.790	0.489	0.490	0.239	0.735	0.386	0.640	0.628	0.550
33631	CAD_2	0.696	0.541	0.457	0.228	0.620	0.260	0.551	0.547	0.487
<u>30772</u>	<u>CAD_3</u>	0.635	0.790	0.860	0.907	0.652	0.852	0.756	0.493	<u>0.743</u>

**Table S2.** Candidate coniferyl alcohol biosynthetic genes from *Sinopodophyllum hexandrum* and their closest characterized homolog in the model plant *Arabidopsis thaliana* according to The *Arabidopsis* Information Resource (TAIR).

	Number of Amino Acids	TAIR BLASTX Top Hit	Num. AAs in TAIR BLASTX Top Hit	TAIR BLASTX Top Hit %ID, %positive
<b>Putative <i>ShPAL</i></b>	716	AT2G37040.1, PAL1	725	82, 91
<b>Putative <i>ShC4H</i></b>	504	AT2G30490.1, C4H	505	77, 86
<b>Putative <i>Sh4CL</i></b>	542	AT1G516801, 4CL1	561	73, 86
<b>Putative <i>ShHCT</i></b>	431	AT5G48930.1, HCT	433	76, 84
<b>Putative <i>ShC3H</i></b>	508	AT2G40890.1, CYP98A3 (C3H)	508	79, 89
<b>Putative <i>ShCCoA-OMT</i></b>	248	AT4G34050.3, CCoAOMT1	286	85, 95
<b>Putative <i>ShCCR</i></b>	337	AT1G15950.3, CCR1	344	66, 75
<b>Putative <i>ShCAD</i></b>	357	AT4G34230.1, CAD5	357	69, 80

**Table S3.** Results of large-scale DPT (A) and (-)-morelensin (B) production and isolation using optimized conditions.

**A.**

(-)deoxypodophyllotoxin (DPT)

Trial	Number of Days from Infiltration to Harvest	Number of Plants and Leaves	Total Mass of Dried Leaves (g)	Theoretical Yield based on LC-MS (mg)	Mass of Isolate from Preparative HPLC (mg)	Pure Yield (mg/g DW) and Percent Yield
1	7	15, 72	2.2421	did not record	1.6	0.71, -
2	9	15, 72	1.5004	did not record	0.9	0.60, -
3	9	15, 72	2.2481	2.82	1.2	0.53, 43%
4	9	20, 75	3.9302	2.18	1.8	0.46, 83%

**B.**

(-)morelensin

Trial	Number of Days from Infiltration to Harvest	Number of Plants and Leaves	Total Mass of Dried Leaves (g)	Theoretical Yield based on LC-MS (mg)	Mass of Isolate from Preparative HPLC (mg)	Pure Yield (mg/g DW) and Percent Yield
1	8	20, 64	2.4858	1.6	1.0	0.40, 63%
2	9	23, 70	2.1206	did not record	2.9	1.4, -

**Table S4.** Candidate laccases from *Sinopodophyllum hexandrum* and their closest characterized homolog in the model plant *Arabidopsis thaliana* according to The *Arabidopsis* Information Resource (TAIR).

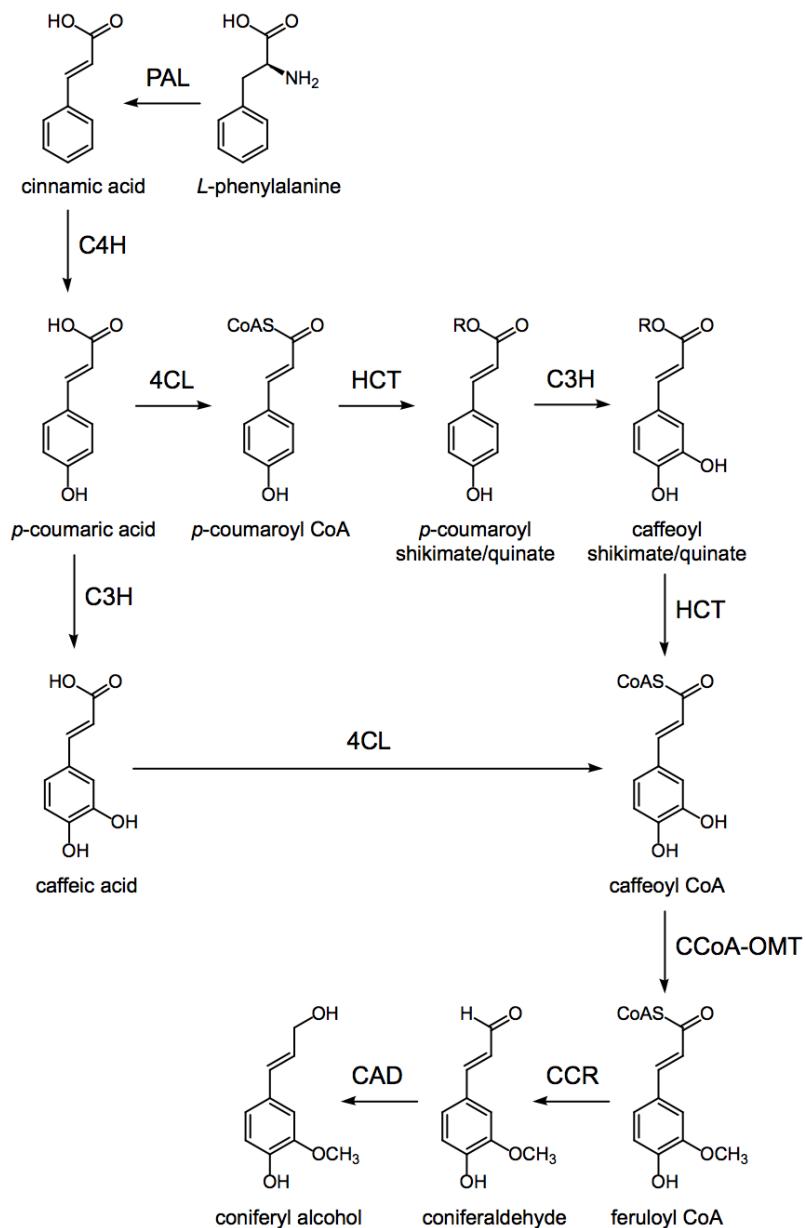
	Number of Amino Acids	TAIR BLASTX Top Hit	Num. AAs in TAIR BLASTX Top Hit	TAIR BLASTX Top Hit %ID, %positive
<b>Putative <i>ShLAC1</i></b>	573	AT5G60020.1, LAC17	577	71, 83
<b>Putative <i>ShLAC2</i></b>	576	AT5G60020.1, LAC17	577	66, 77

**Table S5.** Primers for cloning candidate *S. hexandrum* laccases and coniferyl alcohol biosynthetic genes from *S. hexandrum* cDNA. Capitalized part of sequence indicates part of primer overlapping with the candidate gene sequence. Lowercase part indicates part of primer overlapping with the plasmid sequence. The six letter lowercase parts spaced separately from the longer parts indicate the cut sites of the restriction enzymes used.

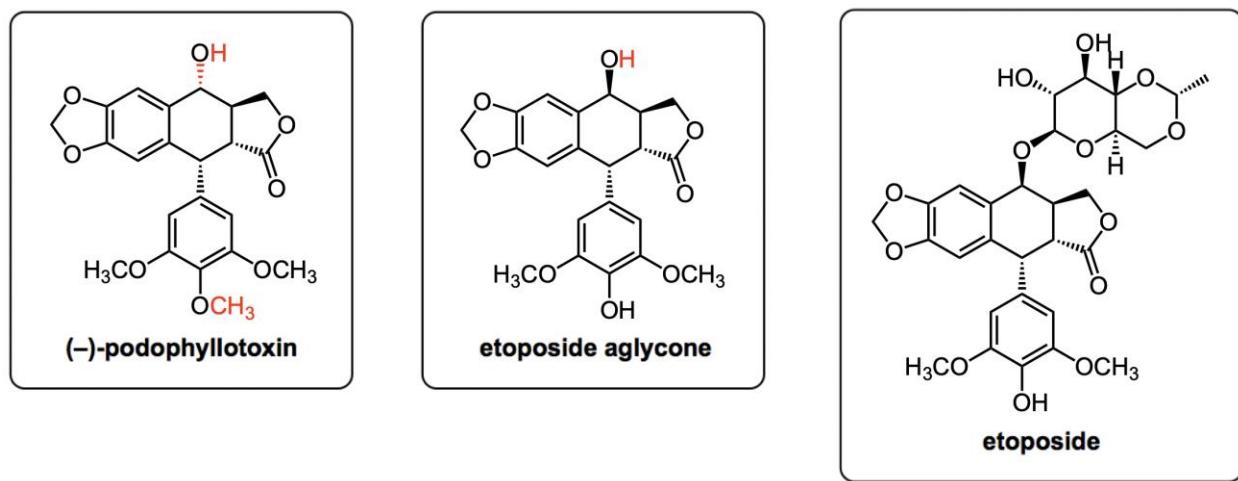
	Forward Primer Sequence	Reverse Primer Sequence
<b>Putative <i>ShLAC1</i></b>	attctgcccaaattcgcg accgg ATGGGTTCACTTCCAAGG	gaaaccagagttaaaggc ctcgag TTAGCACTGGGAAGATCAGACGG
<b>Putative <i>ShLAC2</i></b>	attctgcccaaattcgcg accgg ATGGCATCTCTTGTTTCCCTCAAGG	gaaaccagagttaaaggc ctcgag TTAGCACTTGGGAAGATCAGATGG
<b>Putative <i>ShPAL</i></b>	attctgcccaaattcgcg accgg ATGGGAAGTTAGTGCAAGAAAAC	gaaaccagagttaaaggc ctcgag CTAACAGATAGGAAGAGGGCTACC
<b>Putative <i>ShC4H</i></b>	attctgcccaaattcgcg accgg ATGGATCTCCTCATCTTAGAAAAACTCTC	gaaaccagagttaaaggc ctcgag TCAAAACACTCTGGGTTAACAAAC
<b>Putative <i>Sh4CL</i></b>	attctgcccaaattcgcg accgg ATGGAGACTCCTCCCCAAGAGTACATATT	gaaaccagagttaaaggc ctcgag CTAATTAGGAAGACCAGCTGCTAG
<b>Putative <i>ShHCT</i></b>	attctgcccaaattcgcg accgg ATGATTATTAACGTGAAGGAATCAAC	gaaaccagagttaaaggc ctcgag CTAAAACTCATACAAAATTTTTC
<b>Putative <i>ShC3H</i></b>	attctgcccaaattcgcg accgg ATGGCTCTTCGTGTTCTCCTC	gaaaccagagttaaaggc ctcgag TCACATGTCCACGCCATGCCTTGATAG
<b>Putative <i>ShCCoA-OMT</i></b>	attctgcccaaattcgcg accgg ATGGCAACCAACCAGAC	gaaaccagagttaaaggc ctcgag TCAGCTGAGACGACGGCAGATG
<b>Putative <i>ShCCR</i></b>	attctgcccaaattcgcg accgg ATGCCTGTTGACAGTTGTTACCTC	gaaaccagagttaaaggc ctcgag TCAAGACTGAATGCGGAGGGATTC
<b>Putative <i>ShCAD</i></b>	attctgcccaaattcgcg accgg ATGGGTAGTATTGAAGTTGATAGAAC	gaaaccagagttaaaggc ctcgag CTAACGCATCAAGCTTGCTTCCAGC

**Scheme S1.** Canonical coniferyl alcohol biosynthetic pathway. The shunt pathway that skips the activity of HCT, a shikimate/quinate transferase, is thought to occur to some extent, but its biological relevance remains unclear.

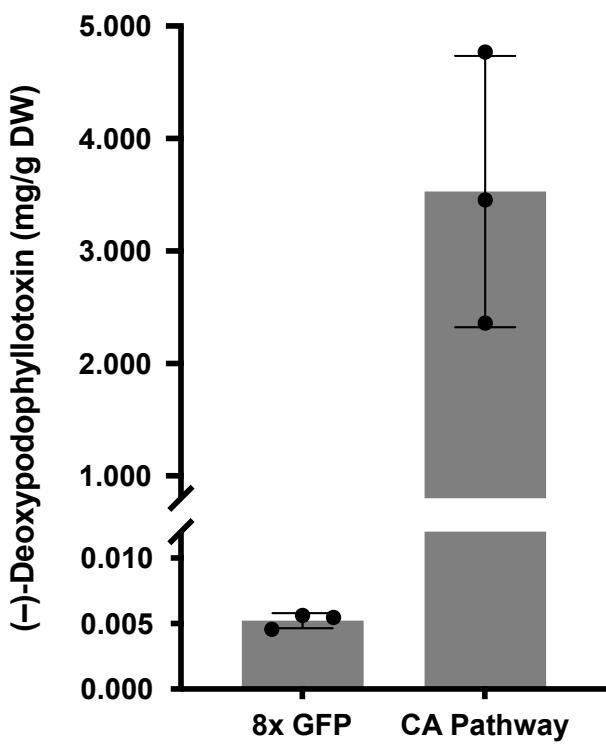
Abbreviations: PAL: phenylalanine ammonia-lyase; C4H: cinnamate 4-hydroxylase; 4CL: 4-coumarate:CoA ligase; HCT: hydroxycinnamoyl-CoA shikimate/quinate hydroxycinnamoyl transferase; C3H: coumarate 3-hydroxylase; CCoA-OMT: caffeoyl CoA O-methyltransferase; CCR: cinnamoyl CoA reductase; CAD: cinnamyl alcohol dehydrogenase.



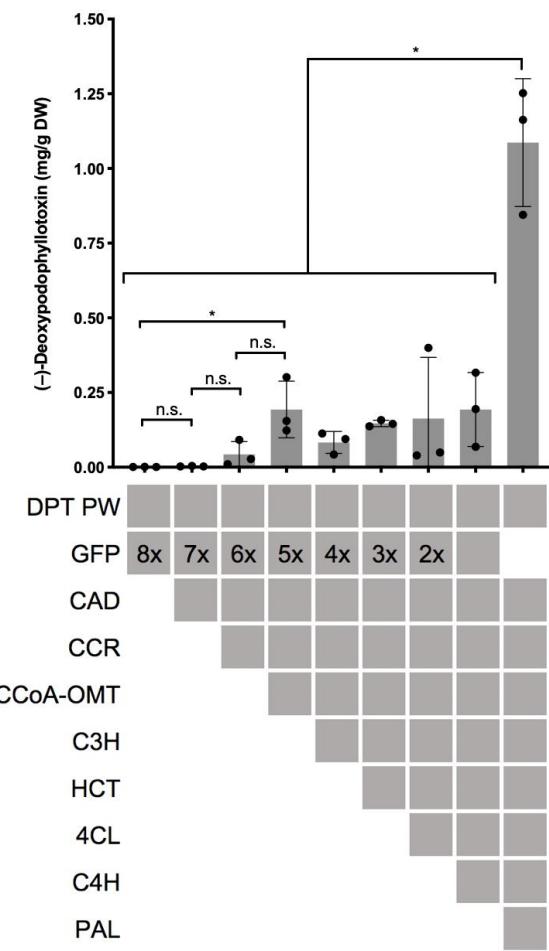
## Supplementary Figures



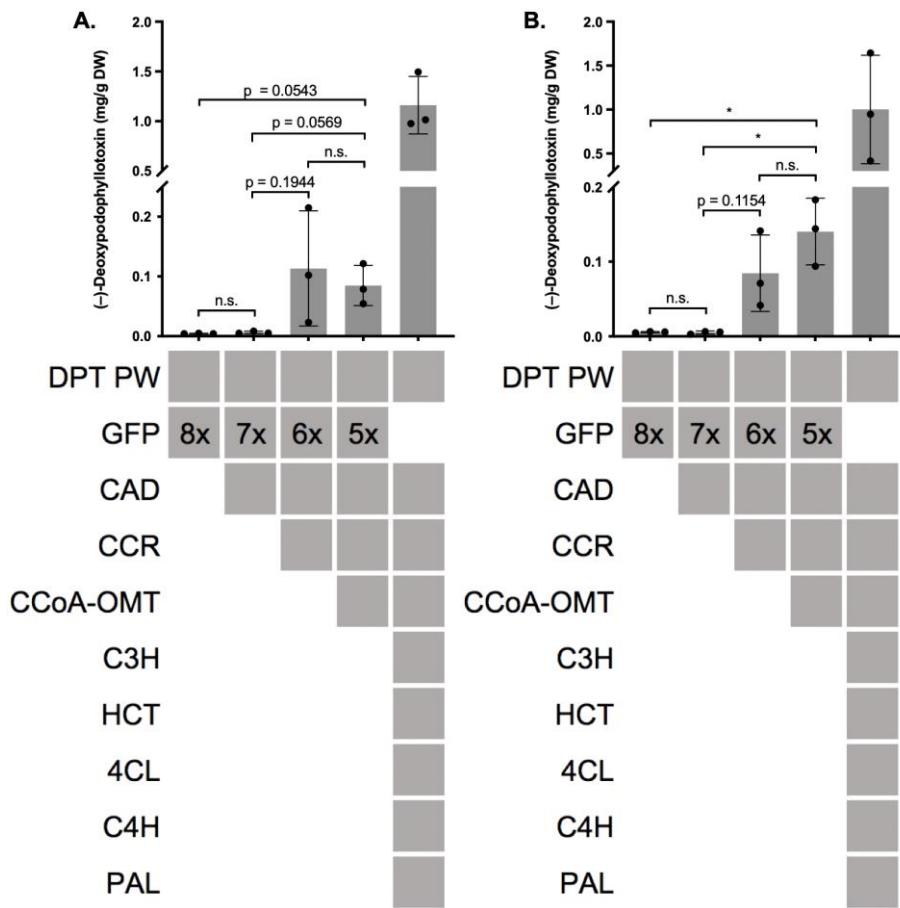
**Figure S1.** Chemical structures of (-)-podophyllotoxin, the current semi-synthetic precursor to etoposide; the etoposide aglycone, a molecule that can be accessed biosynthetically and that is synthetically closer to etoposide due to its demethylated lower ring oxygen and epimerized upper core hydroxyl group; and etoposide itself.



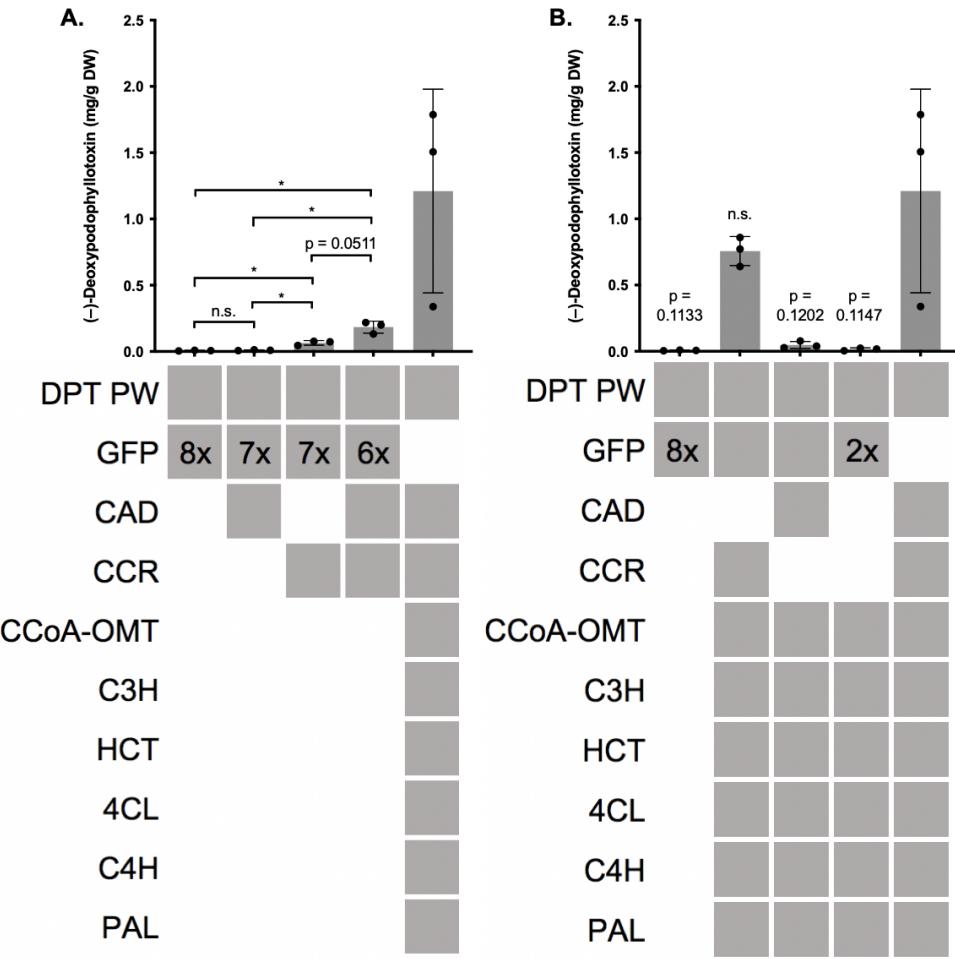
**Figure S2.** Production of DPT in *N. benthamiana* plants transiently expressing the DPT pathway genes in addition to either the eight coniferyl alcohol biosynthetic genes (right) or an equal total level of GFP according to the OD of the *A. tumefaciens* strains (left). Bars depict mean ( $\pm$  SD) and dots depict individual biological replicates for each set of conditions.



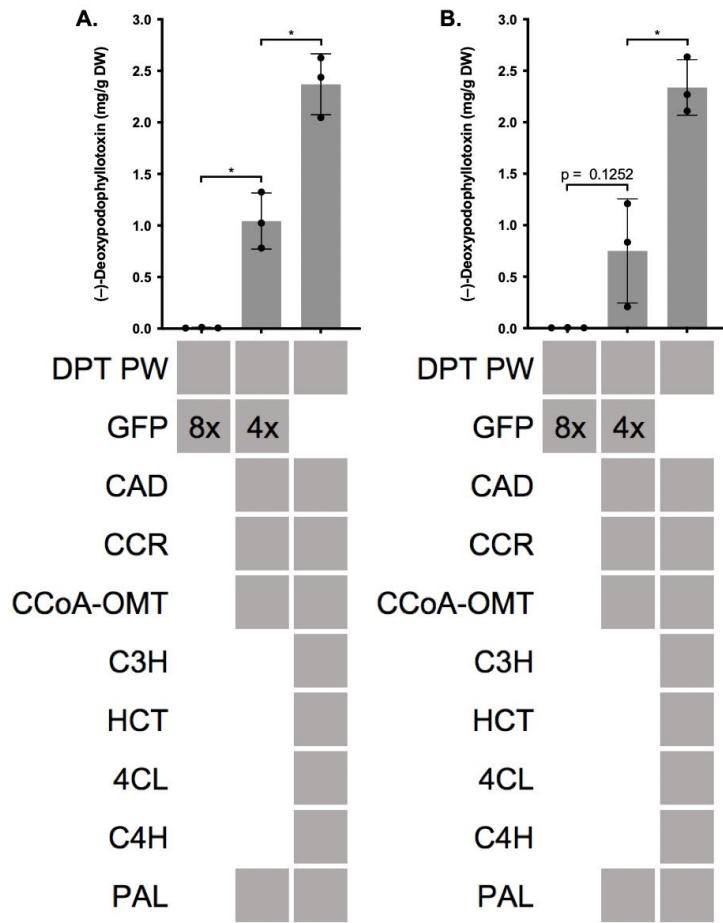
**Figure S3.** Replicate of the experiment shown in Figure 2 with a different batch of plants. Bars depict mean ( $\pm$  SD) and dots depict individual biological replicates for each set of conditions. Grey boxes indicate presence of the given *Agro.* strain in the mixture, while “#x” labels show the relative OD proportion of the relevant strain compared to each individual strain in the DPT pathway. Asterisks indicate a significant difference between sets, with  $p < 0.05$  according to Welch’s t-test. Certain set pairings with no significant difference are denoted with “n.s.” Data are shown here to demonstrate the variation present in these types of experiments: while the addition of PAL with the rest of the coniferyl alcohol biosynthetic enzymes clearly results in a very large, significant increase compared to the GFP-only control, the contributions of some of the other individual enzymes in the pathway is not as clear.



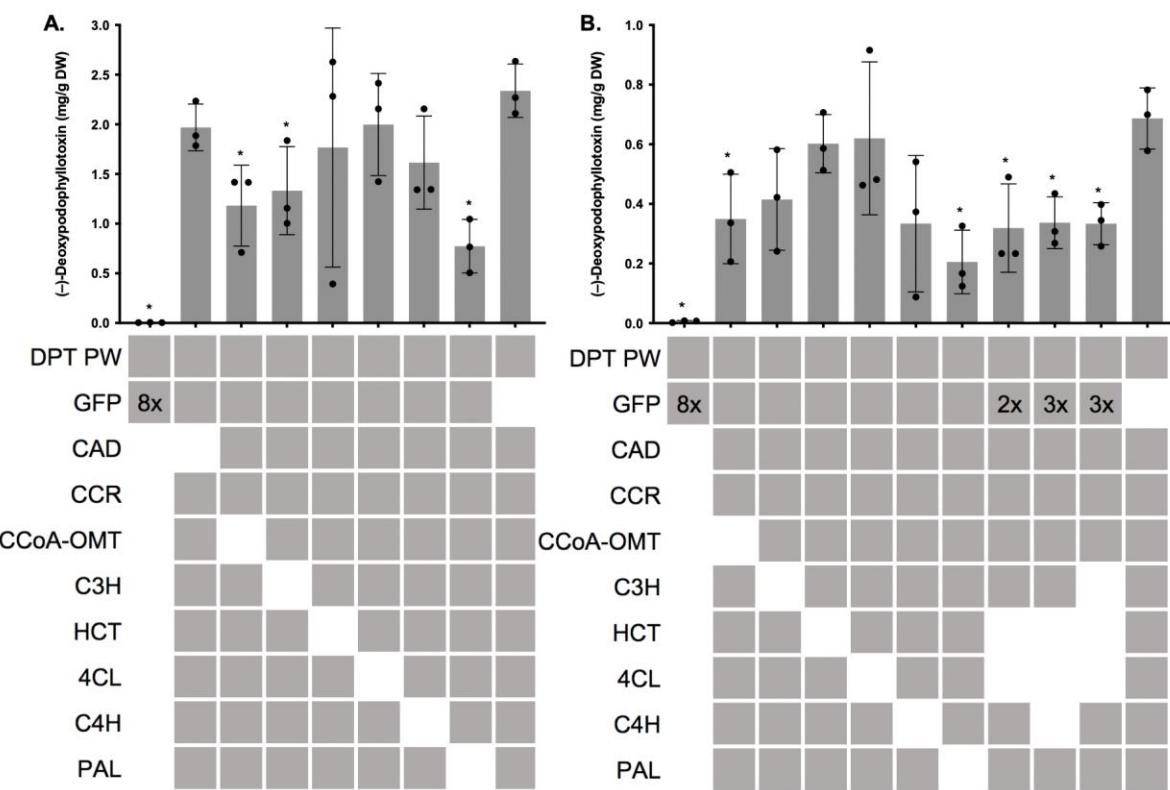
**Figure S4.** Determination of the individual roles of CAD, CCR, and CCoA-OMT in DPT production via sequential addition. (A) and (B) show two replicates of the same experiment with different batches of plants. Bars depict mean ( $\pm$  SD) and dots depict individual biological replicates for each set of conditions. Asterisks indicate a significant difference between sets, with  $p < 0.05$  according to Welch's t-test. Some other low p-values are shown for comparison purposes. P-values greater than 0.2 are marked with "n.s." for "not significant." PW = pathway.



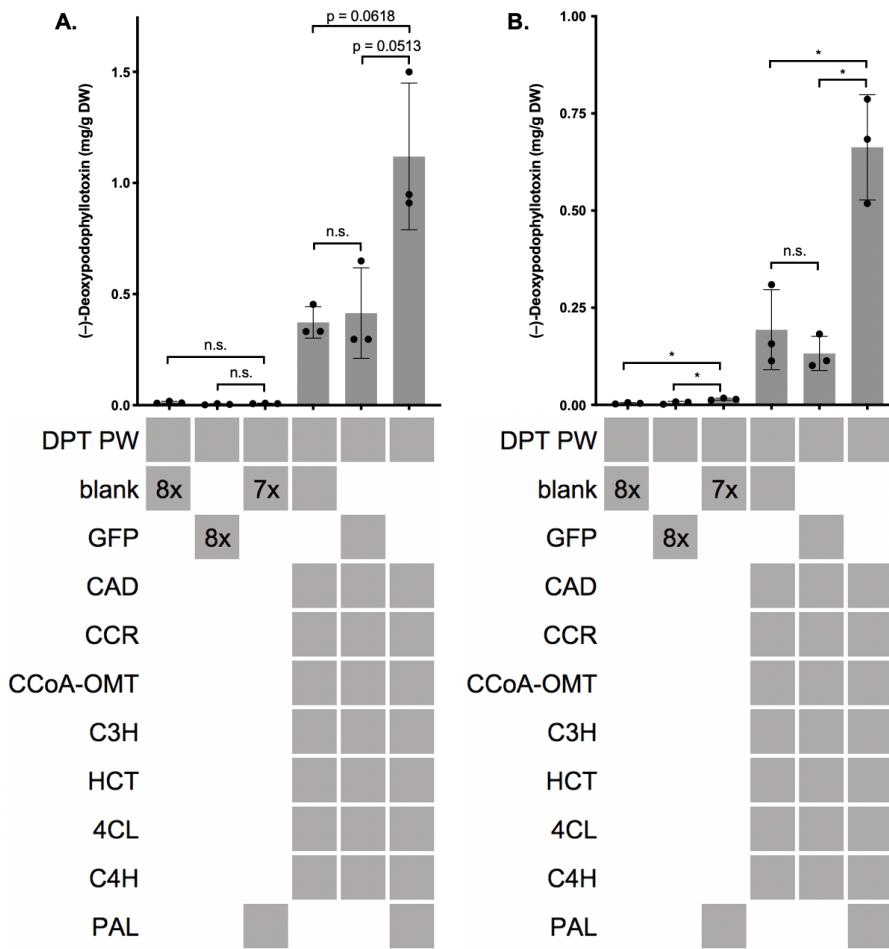
**Figure S5.** Determination of the roles of CAD and CCR in DPT production. Bars depict mean ( $\pm$  SD) and dots depict individual biological replicates for each set of conditions. Asterisks indicate a significant difference between sets, with  $p < 0.05$  according to Welch's t-test. Some other low p-values are shown for comparison purposes. (A) shows an experiment in which the impact of these two genes was tested in isolation and combination, while (B) shows an experiment in which the impact of the lack of either or both of these genes was tested with the inclusion of the other coniferyl alcohol biosynthetic genes. All significance tests in B are comparisons with the right-most set (DPT pathway + full CA pathway). PW = pathway.



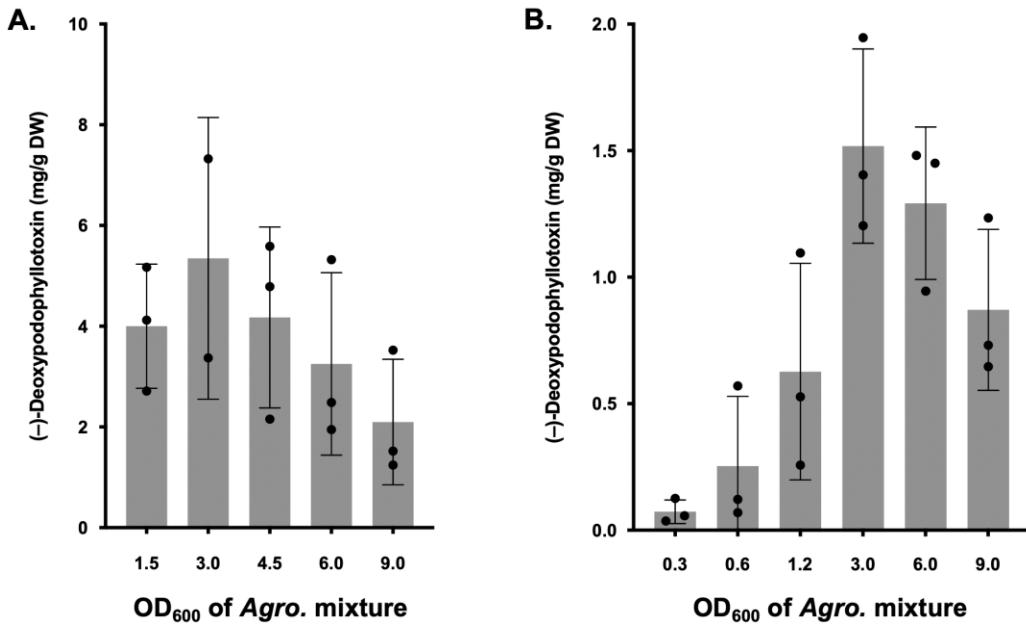
**Figure S6.** Determination of the collective role of C3H, HCT, 4CL, and C4H in DPT production. (A) and (B) show two replicates of the same experiment with different batches of plants. Expression of these four enzymes does not appear to contribute significantly to flux according to the “stepwise addition” experiments of Figures 2 and S3; hence they were tested here together to determine if they could be left out completely without decreasing DPT production. Bars depict mean ( $\pm$  SD) and dots depict individual biological replicates for each set of conditions. Asterisks indicate a significant difference between sets, with  $p < 0.05$  according to Welch’s t-test. Another (low) p-value is shown for comparison purposes. PW = pathway.



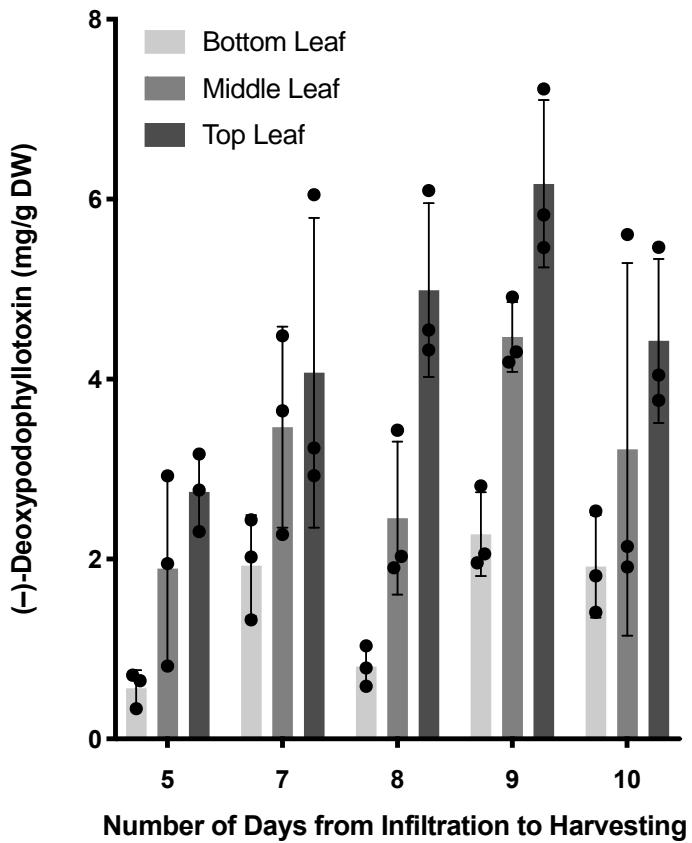
**Figure S7.** Determination of the roles of CCoA-OMT, C3H, HCT, 4CL, C4H, and PAL in DPT production. Bars depict mean ( $\pm$  SD) and dots depict individual biological replicates for each set of conditions. (A) includes a “drop-out” experiment for CAD, while (B) includes combination “drop-out” experiments with C3H, HCT, 4CL, and C4H. Asterisks in both panels indicate a significant difference between the given set and the full coniferyl alcohol pathway set (farthest right in each panel), with  $p < 0.05$  according to Welch’s t-test. PW = pathway.



**Figure S8.** Impact of PAL expression on DPT production with and without GFP expression. (A) and (B) show two replicates of the same experiment with different batches of plants, in which the goal was to ensure that lack of GFP expression was not the reason for the increase in yields rather than inclusion of PAL expression. Here, “blank” refers to a volume of *Agrobacterium* induction media added to the infiltration mixture in place of a volume of GFP-harboring *A. tumefaciens* strain. Bars depict mean ( $\pm$  SD) and dots depict individual biological replicates for each set of conditions. Asterisks indicate a significant difference between sets, with  $p < 0.05$  according to Welch’s t-test. Some other low p-values are shown for comparison purposes. PW = pathway.

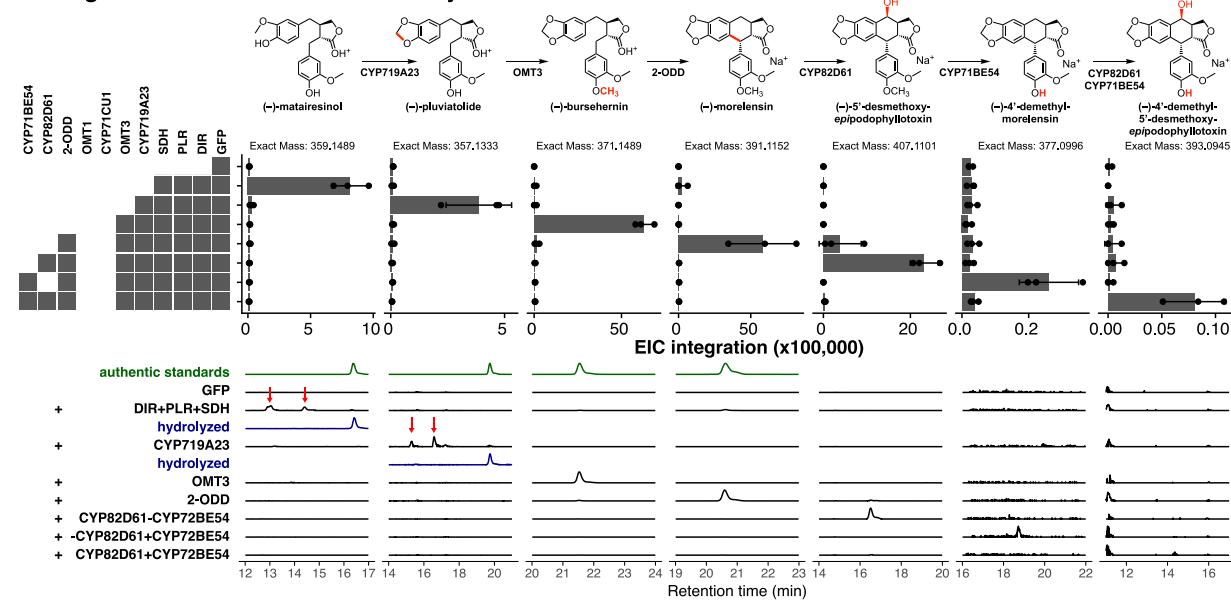


**Figure S9.** Impact of total *A. tumefaciens* strain mixture OD<sub>600</sub> on DPT yield. (A) and (B) show two variations of the same experiment (done months apart). All experiments shown include the full DPT and coniferyl alcohol biosynthetic pathways. Bars depict mean ( $\pm$  SD) and dots depict individual biological replicates for each set of conditions.

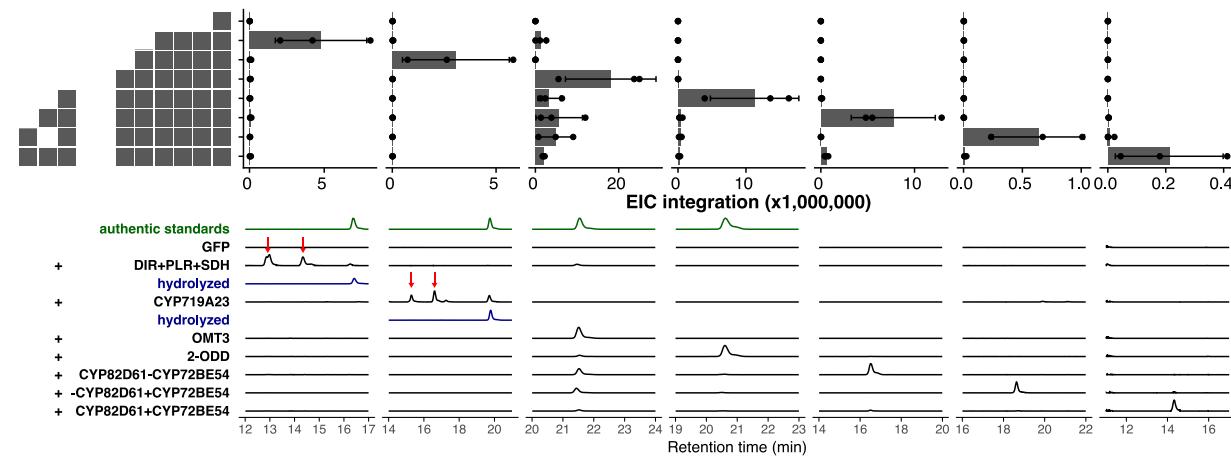


**Figure S10.** Impact of infiltration-to-harvest time on DPT accumulation/production. Bars depict mean ( $\pm$  SD) and dots depict individual biological replicates for each set of conditions. Each cluster of three bars represents three plants total. “Top” vs. “middle” vs. “bottom” leaf indicates relative age of each leaf at the time of infiltration (all infiltrations occurred at the same time). “Top” leaves are newer, while leaves lower on the plant stem (“bottom”) are older.

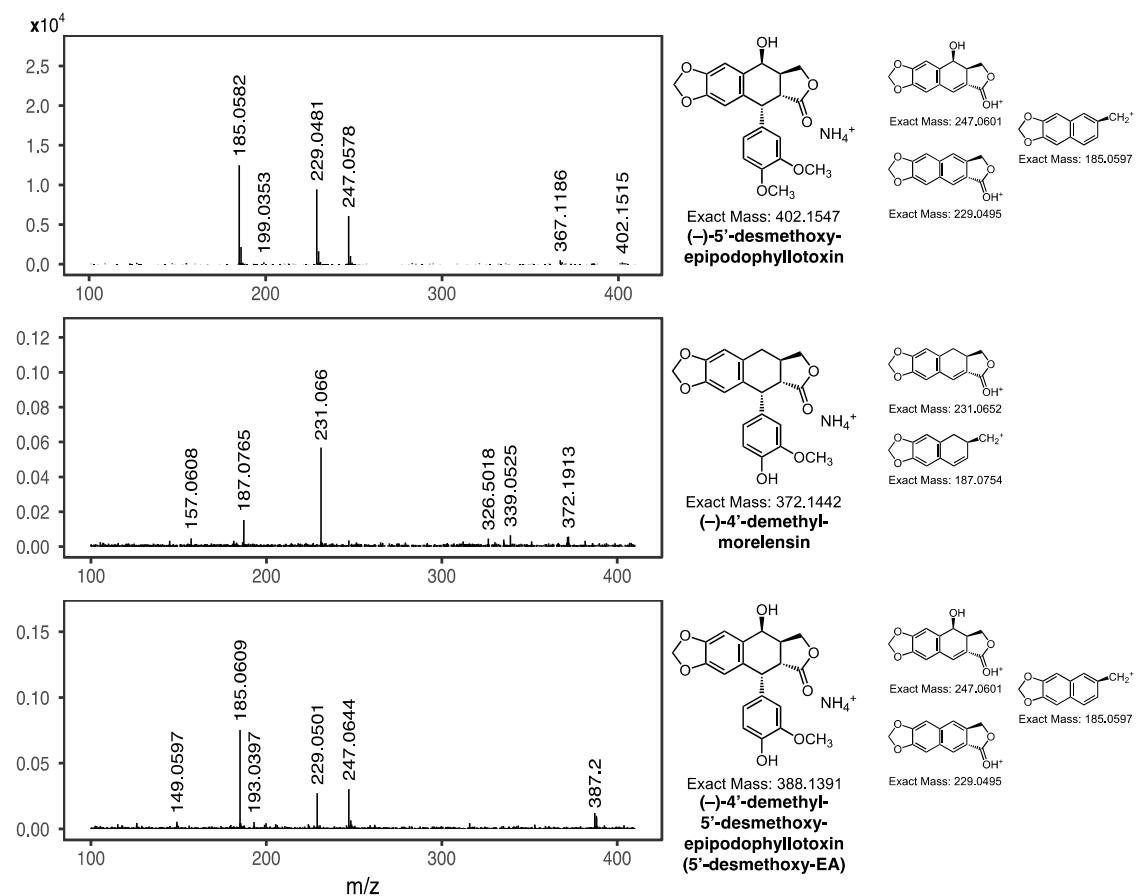
### A. Exogenous addition of 1 mM coniferyl alcohol



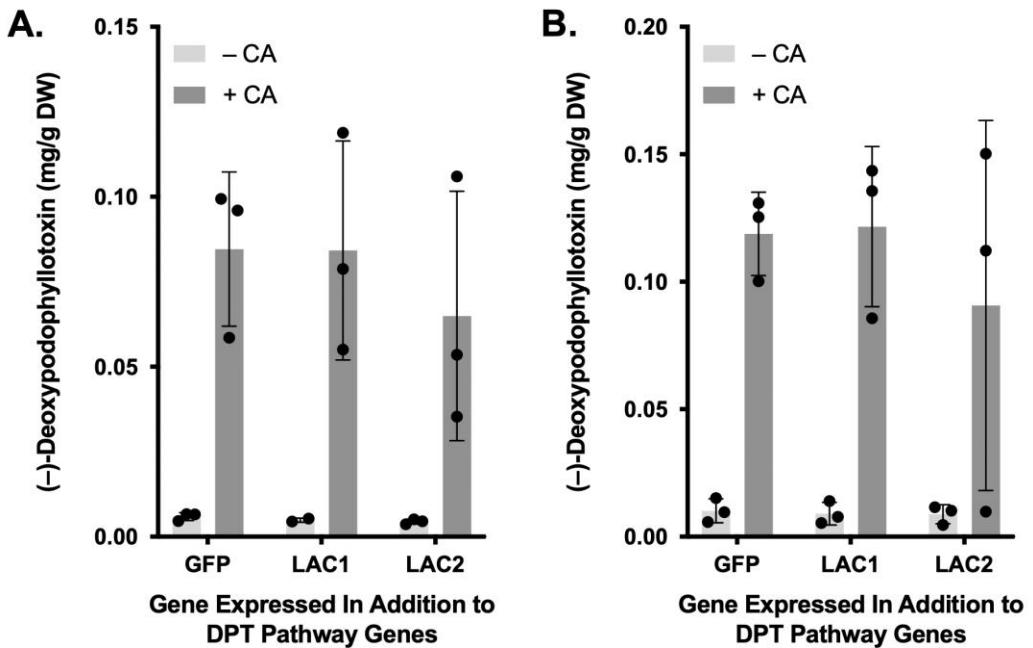
### B. Co-expression of coniferyl alcohol pathway enzymes



**Figure S11.** Production of (-)-4'-demethyl-5'-desmethoxy-epipodophyllotoxin (5'-desmethoxy-EA) and its precursors in *N. benthamiana* detected by LC-MS with (A) exogenous coniferyl alcohol addition (1 mM) or (B) CA pathway expression. With co-expression of CA pathway enzymes and with any combination of EA pathway genes including 2-ODD, (-)-bursehernin accumulation is observed from incomplete processing by 2-ODD, likely due to an order of magnitude increase of the precursor, (-)-bursehernin, supply. As denoted in the chemical structures drawn above, (-)-matairesinol, (-)-pluviatolide, and (-)-bursehernin were detected as  $[M+H]^+$ , and (-)-morelensin, (-)-5'-desmethoxy-epipodophyllotoxin, (-)-4'-demethyl-morelensin and 5'-desmethoxy-EA as  $[M+Na]^+$ . Red arrows indicate in-source fragmentation products of the glycosylated derivatives of corresponding compounds. Acid hydrolysis treatment allowed recovery of the parent compounds compared against authentic standards (blue traces: acid-hydrolyzed plant extracts; green traces: authentic standards; see Figure S12 for putative structures for the last three compounds). Data bars indicate the mean peak integration of corresponding extracted ion chromatogram (EIC) traces, and error bars standard deviations from three biological replicates. Gray-shaded boxes indicate enzymes expressed in combination via Agro.-infiltration.



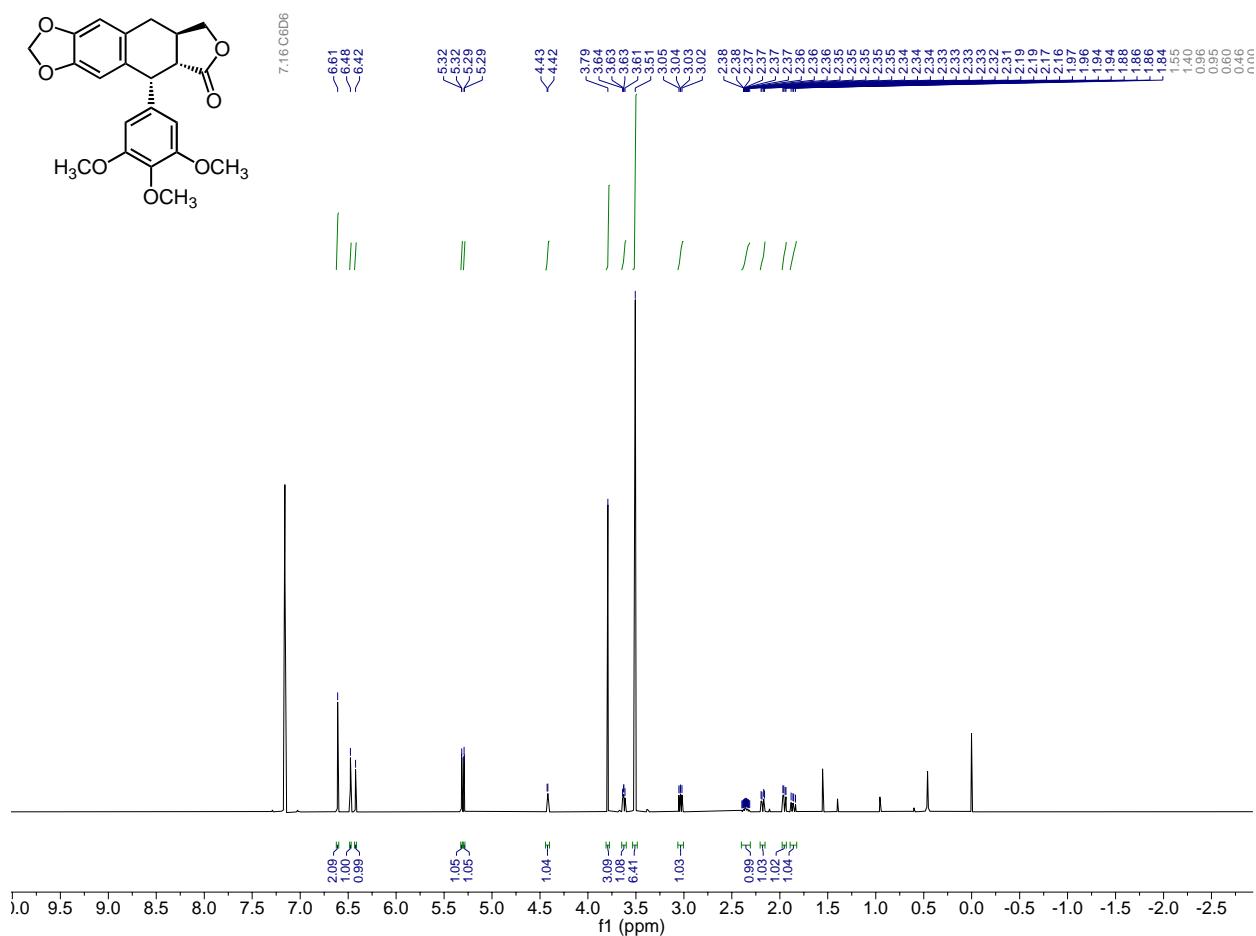
**Figure S12.** MS/MS spectra of putative (-)-5'-desmethoxy-epipodophyllotoxin, (-)-4'-demethyl-morelensin, and 5'-desmethoxy-EA, and proposed structures for the fragmentation patterns.



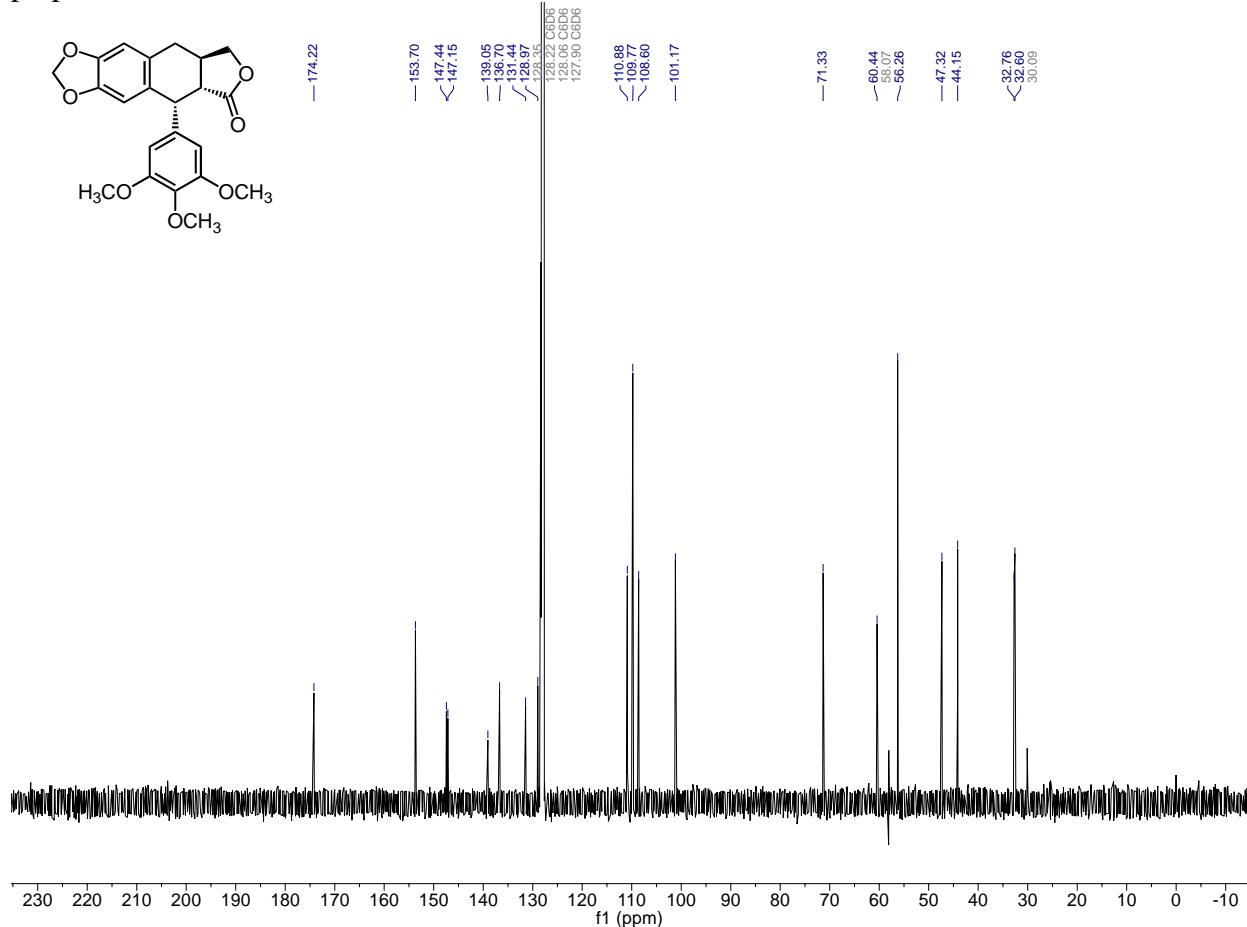
**Figure S13.** Impact of laccase candidate expression on DPT production. (A) and (B) show results from two replicates of the same experiment with different batches of plants. Bars depict mean ( $\pm$  SD) and dots depict individual biological replicates for each set of conditions. Color indicates absence (-CA) or presence (+CA) of coniferyl alcohol infiltration. No significant differences in DPT production between expression of laccase candidates and GFP were observed.

## Supplementary Spectra

**Spectrum 1.**  $^1\text{H}$ -NMR of isolated (-)-deoxypodophyllotoxin from *N. benthamiana* leaves during large-scale trial (20 plants, 75 leaves). Post-silica gel column flash purification and post-preparative-HPLC.

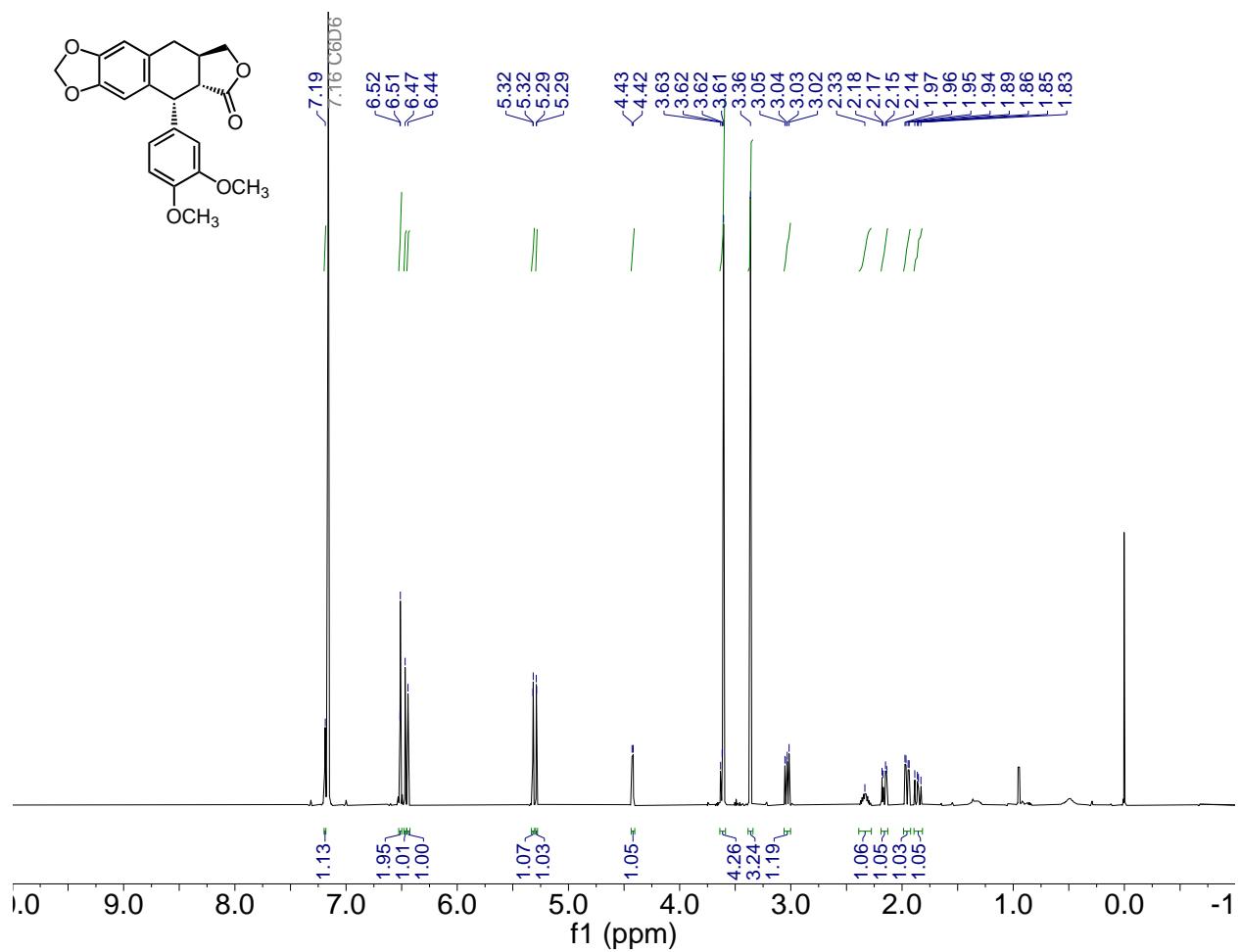


**Spectrum 2.**  $^{13}\text{C}$ -NMR of isolated (-)-deoxypodophyllotoxin from *N. benthamiana* leaves during large-scale trial (20 plants, 75 leaves). Post-silica gel column flash purification and post-preparative-HPLC.

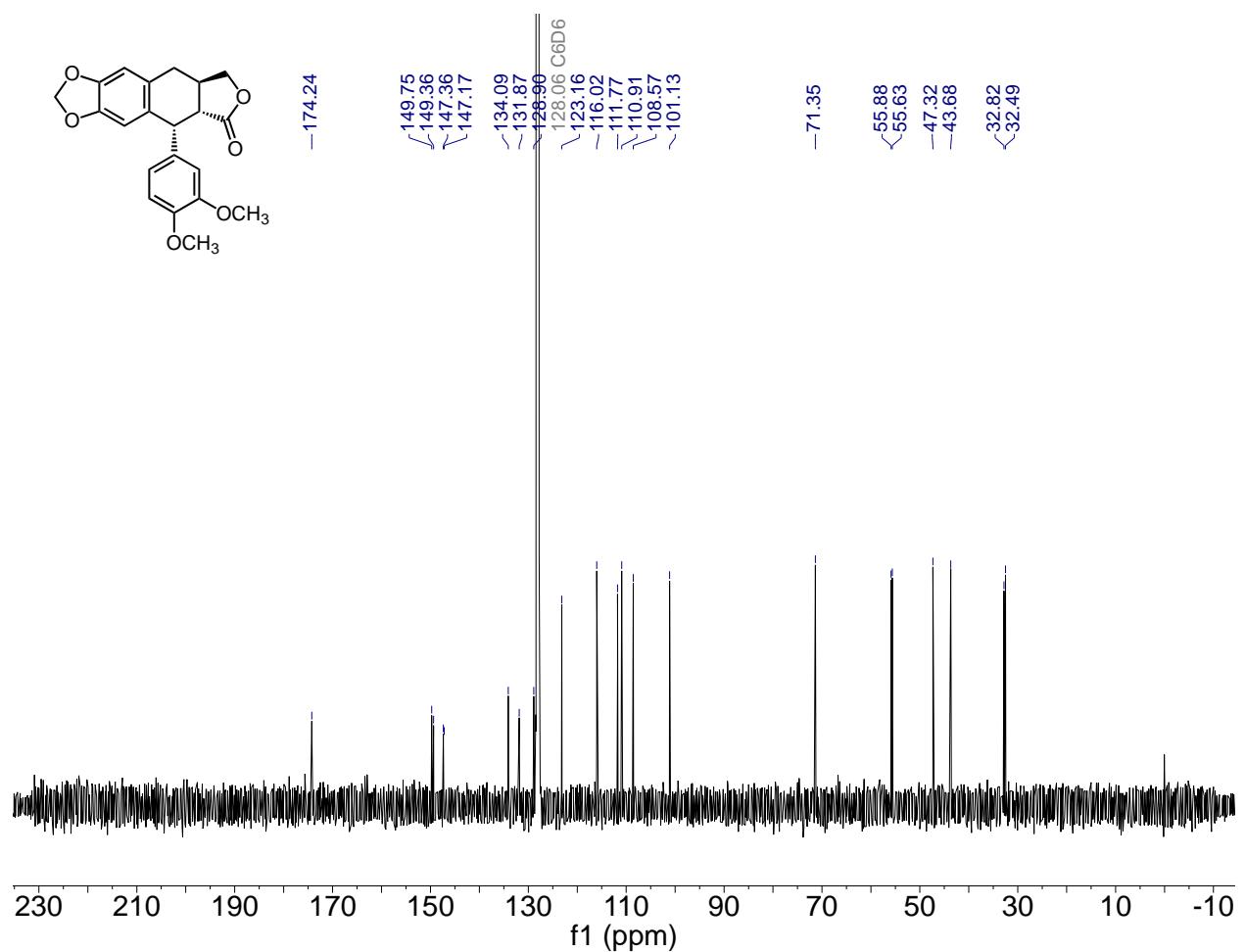


Experimental  $^{13}\text{C}$  δ ( $\text{C}_6\text{D}_6$ , 151 Hz): 174.2 (C-9'), 153.6 (C-3', C-5'), 147.4 (C-5), 147.1 (C-4), 139.0 (C-4'), 136.6 (C-1'), 131.4 (C-2), 128.9 (C-1), 110.8 (C-3), 109.7 (C-6), 108.5 (C-6', C-2'), 101.1 (O- $\text{CH}_2$ -O), 71.3 (C-9), 60.4 (CH<sub>3</sub>O-4'), 56.2 (CH<sub>3</sub>O-3', CH<sub>3</sub>O-5'), 47.3 (C-8), 47.3 (C-8'), 44.1 (C-7'), 32.5 (C-7)

**Spectrum 3.**  $^1\text{H}$ -NMR of isolated (-)-morelensin from *N. benthamiana*.



**Spectrum 4.**  $^{13}\text{C}$ -NMR of isolated (-)-morelensin from *N. benthamiana* leaves.



## Supplementary References

- (1) Lau, W.; Sattely, E. S. Six enzymes from mayapple complete the biosynthetic pathway to the etoposide aglycone. *Science* **2015**, *349*, 1224-1228.
- (2) Sainsbury, F.; Thuenemann, E. C.; Lomonossoff, G. P. pEAQ: versatile expression vectors for easy and quick transient expression of heterologous proteins in plants. *Plant Biotechnol. J.* **2009**, *7*, 682-693.
- (3) Gibson, D. G.; Young, L.; Chuang, R.-Y.; Venter, J. C.; Hutchison III, C. A.; Smith, H. O. Enzymatic assembly of DNA molecules up to several hundred kilobases. *Nat. Methods* **2009**, *6*, 343-345.
- (4) Thuring, R. W. J.; Sanders, J. P. M.; Borst, P. A freeze-squeeze method for recovering long DNA from agarose gels. *Anal. Biochem.* **1975**, *66*, 213-220.
- (5) Holsters, M.; de Waele, D.; Depicker, A.; Messens, E.; van Montagu, M.; Schell, J. Transfection and Transformation of *Agrobacterium tumefaciens*. *Mol. Gen. Genet.* **1978**, *163*, 181-187.
- (6) Eudes, A.; Sathitsuksanoh, N.; Baidoo, E. E.; George, A.; Liang, Y.; Yang, F.; Singh, S.; Keasling, J. D.; Simmons, B. A.; Loqué, D. Expression of a bacterial 3-dehydroshikimate dehydratase reduces lignin content and improves biomass saccharification efficiency. *Plant Biotechnol. J.* **2015**, *13*, 1241-1250.
- (7) Rojas-Sepúlveda, A. M.; Mendieta-Serrano, M.; Antúnez Mojia, M. Y.; Salas-Vidal, E.; Marquina, S.; Villarreal, M. L.; Puebla, A. M.; Delgado, J. I.; Alvarez, L. Cytotoxic Podophyllotoxin Type-Lignans from the Steam Bark of *Bursera fagaroides* var. *fagaroides*. *Molecules* **2012**, *17*, 9506-9519.
- (8) Cho, E. J.; Choi, J. M.; Kim, H. M.; Choi, K.; Ku, J.; Park, K.-W.; Kim, J.; Lee, S. Antibacterial Activity and Protective Effect against Gastric Cancer by *Anthriscus sylvestris* Fractions. *Hortic., Environ. Biotechnol.* **2013**, *54* (4), 326-330.
- (9) Jolad, S. D.; Wiedhopf, R. M.; Cole, J. R. Cytotoxic Agents from *Bursera morelensis* (Burseraceae): Deoxypodophyllotoxin and a New Lignan, 5'-Desmethoxydeoxypodophyllotoxin. *J. Pharm. Sci. (Philadelphia, PA, U.S.)* **1977**, *66* (6), 892-893.
- (10) Ikeda, R.; Nagao, T.; Okabe, H.; Nakano, Y.; Matsunaga, H.; Katano, M.; Mori, M. Antiproliferative Constituents in Umbelliferae Plants. IV.<sup>1)</sup> Constituents in the Fruits of *Anthriscus sylvestris* Hoffm. *Chem. Pharm. Bull.* **1998**, *46* (5), 875-878.